

Due process or French farce?

France's 'contaminated blood' trial is a reminder that witch-hunts, however well disguised in legal formalities, are no substitute for a credible, wide-ranging and dispassionate judicial inquiry.

Britain's judicial inquiry into the causes of its bovine spongiform encephalopathy (BSE) crisis may be belated, but it is at least being carried out thoroughly, rigorously and dispassionately (see *Nature* **392**, 532; 1998). The contrast with France's long-running contaminated blood affair could not be more stark.

In the latest round, three former ministers, including former prime minister Laurent Fabius, went on trial in Paris last week charged with 'involuntary homicide' for their handling of the threat of HIV infection of the blood supply in 1985 (see page 548). They are accused in particular of delaying the introduction of a US AIDS test — to screen blood donations — to protect the market for a French test.

The trial has quickly turned to farce. Verdicts will be delivered by a group of parliamentary judges in the Court of Justice of the Republic, a court created specially to try the ministers. But so far, instead of proceeding with a detailed blow-by-blow analysis of the chronology of the affair and the many complex issues involved, the court has engaged in *ad hoc* questioning of the defendants and witnesses. Matters have not been helped by the performance of the president, who visibly lacks an adequate grasp of the dossier.

The resulting dire lack of serious cross-examination has rightly prompted protests from victims infected with HIV via transfusions, and their families, who feel that justice may not be done. Justice for the defendants may also fail to emerge: even if judged innocent, the questionable legal procedures mean that public suspicion of their guilt will hang over them for the rest of their days.

This sad state of affairs was fairly predictable. The government, its advisory committees and medical institutions have been justifiably forced by the media and the judiciary to account for their actions. But the subsequent criminalization of the whole issue has served no one.

Confusing things further is the fact that what is on trial is not simply the actions of individuals that led to the blood scandal, but the entire system of a ruling élite and its advisers, usually graduates of the prestigious *grandes écoles* and considered removed from the interests of the people.

Also, as so often in this affair, the media are conducting their own fringe trial in the corridors of the court. The atmosphere remains so emotionally charged that cold scientific arguments by the defence, no matter how sound, cannot hope to compete with the impact of a televised interview with a mother who has lost her child to AIDS through a blood transfusion. Reason and science have gone out of the window, and the cherished principles of the serenity of justice and presumption of innocence have been trampled upon.

Scientists who have testified to the extenuating circumstances of the mid-1980s have too often been summarily rejected as merely bidding to absolve the guilt of those in positions of responsibility at the time. Michel Setbon, a French researcher, told the court last week that his research showed that "there is no relationship of cause and effect between the introduction of the diagnostic tests and the [level] of contaminations".

If he is right, the prosecution's case may already be in tatters. More enlightening, however, would have been an independent analysis of not just this aspect, but the entire complex affair. But none of the various trials have commissioned independent expert reports, relying instead on documents often removed from their wider context and the sometimes dubious recollections of testifying experts. There is a danger that the saga will yield neither justice nor the lessons that need to be drawn on handling risk and scientific uncertainty if France is to avoid similar tragedies in the future. □

Food for thought

A row over genetically modified foods spotlights the importance of procedures for validating scientific data.

If the stakes were not so high, there would also be something farcical about the past week's media circus in Britain over the potential health risks of genetically modified (GM) crops. Six months ago, Arpad Pusztai, a research scientist at the Rowett Research Institute in Scotland, was temporarily suspended after allegedly misleading the media about the potential health impacts of eating genetically modified foods, based on his (unpublished) studies of rats fed potatoes modified to produce lectin. Now, accompanied by even bigger headlines, a group of his colleagues has endorsed his (still unpublished) conclusions, fanning an already heated debate over a possible moratorium on the commercial growing of all GM crops (see page 547).

The concerns at the centre of the debate are certainly genuine. Given the widespread — and rapidly growing — use of GM products in the food of the average consumer, any indication of a health risk that has not been properly characterized would be a source of major concern. Memories of the BSE crisis, and the continuing reassurances of government officials and scientists that British beef was safe, remain fresh. And, like the nuclear issue, much of the opposition to GM foods rests on legitimate worries that an enthusiasm for a new

technology may have blunted sensibilities to — and encouraged secrecy about — its potential dangers.

But this is a case for intensive monitoring, not a complete ban or even a moratorium. It is also a case for a hard-headed scientific appraisal of evidence and its potential implications, not yet more science by press conference. Philip James, the head of the Rowett Institute — and a respected nutrition researcher — is correct to have reduced the heat of the current debate by opening the available scientific data to public scrutiny. Less understandable are his provocative actions against Pusztai last summer.

There are several lessons to be learnt. One is the reinforcement of the need for the promised Freedom of Information Act; suspicion thrives on a culture of secrecy that is whipped up by stories of scientists being 'gagged'. The media need a better understanding of the difference in credibility between published and unpublished 'scientific' data. And an independent body needs to be set up to assess available (even if preliminary) evidence of the potential dangers of GM foods before full commercialization proceeds. The lessons of BSE have yet to be fully absorbed. □



Gag on food scientist is lifted as gene modification row heats up...

[LONDON] A Scottish research institute this week lifted a ban on one of its scientists from speaking to the press, in a bid to dampen allegations that it had tried to suppress data indicating potential health risks in genetically modified food.

Arpad Pusztai, a senior researcher at the Rowett Research Institute in Aberdeen and an authority on lectins, had been banned from speaking to the press after allegedly disclosing unpublished data from his experiments on television last August.

Pusztai concluded from his data that when rats were fed with potatoes genetically modified to produce a lectin with known insecticidal properties they suffered retarded growth, depressed immune systems and reduced body weight. He was suspended for 12 days during an internal investigation, which exonerated him of any wrongdoing. But his contract was not renewed, and he retired at the end of last year at the age of 68.

Shortly after the row broke out, the institute arranged for an audit of Pusztai's results, which concluded that his data were "too inconsistent" to support his conclusions. But last week, amid mounting media interest and public alarm, more than 20 scientists from 13 countries said in a statement organized by Pusztai's supporters that they had reviewed his work and supported his conclusions.

The scientists accused the institute's director, Philip James, of gagging Pusztai and suppressing the flow of scientific information. They criticized James for refusing to make public an unpublished report by Pusztai justifying his conclusions.

James, the author of the British government's proposals for an independent food standards agency, denies the scientists' allegations that he tried to suppress commercially and politically sensitive scientific results (see *Nature* 394, 714; 1998). He says Pusztai was simply banned from disclosing further unpublished work to the media.

James says the ban is being lifted to counter the view "that Pusztai is subject to some kind of gagging order". James adds that he is planning to ask an independent learned society, such as the Royal Society of Edinburgh, to convene an open, two-day meeting to discuss the issues raised by the affair.

He says he has "reluctantly" made public Pusztai's unpublished report, adding that he would have preferred what has become known as the "alternative report" to have been published first in a peer-reviewed journal. "It is simply not true to say Arpad is gagged," says James. "I offered him a press conference [to publicize his research]. But



At odds: James (left) says that Pusztai was only prevented from discussing his unpublished data.

on condition that he should get it published in a journal first."

Pusztai's supporters claim that the institute's audit report is selective in the data it interprets. "I have the impression that it was hastily compiled and systematically biased towards brushing aside [Pusztai's] experimental findings," says Vyvyan Howard, head of research in fetal and infant toxico-pathology at the University of Liverpool.

Similarly, Stanley Ewen, professor of histopathology at the University of Aberdeen medical school, and a former collaborator with Pusztai, accuses the audit committee of deliberately withholding critical data on the weights of rat organs.

Ewen has been reported as claiming to have confirmed Pusztai's findings from a his-

tological analysis of tissue from Pusztai's rats. Ewen says he is trying to publish this research, so far without success.

But Andrew Chesson, head of the Rowett's nutritional chemistry unit, who chaired the audit committee, says it took account of all the data then available. The panel included an immunologist, and a plant biotechnologist from outside the institute.

Chesson says that, in many cases, the error bars on the data were almost as large as the numerical quantities. "There is simply no way that Pusztai could support his conclusions with these data."

Earlier this week, Pusztai was continuing to refuse to speak publicly. But Luke Anderson, an environmental activist and Pusztai supporter, says the researcher is angry and hurt at how a career in which he produced 280 research papers and wrote three books has been brought to an abrupt end.

Anderson adds that one journal was offered the data, but declined to publish it. He acknowledges that few of the signatories of last week's statement are molecular biologists or immunologists. But he argues that they are all sufficiently qualified in medicine and biology to provide an informed opinion on Pusztai's data. Pusztai's results are expected to be made public this week on <http://www.rri.sari.ac.uk> Ehsan Masood

...and Blair resists demands for a moratorium

[LONDON] The past two weeks have seen unprecedented media coverage in Britain of genetic modification issues. Many of the reports have called for stronger regulation and for a moratorium on the commercial growing of crops until the risks to human health and the environment are better understood.

The government, led by Prime Minister Tony Blair, is continuing to resist calls for a moratorium from opposition political parties and from its own wildlife advisory body, English Nature.

It is also resisting calls from the main opposition Conservative party for the dismissal of the science minister, Lord David Sainsbury, an enthusiastic supporter of biotechnology.

Before becoming a minister, Sainsbury was chairman of the supermarket group that bears his name when it started selling genetically modified food last year. He now chairs an advisory panel on public perceptions of the biosciences.

John Redwood, Conservative spokesman on trade and industry, said: "We need a minister who is independent. David Sainsbury made up his mind years ago and is an advocate of the technology." But Jack Cunningham, secretary of state for the cabinet office, defended Sainsbury as a minister of high personal integrity.

Attention has also focused on whether Britain's regulatory framework is

sufficiently rigorous to ensure that commercial releases of genetically modified crops do not harm human health and the environment.

Derek Burke, former chairman of the Advisory Committee on Novel Foods and Processes, says that any application to commercialize products containing genetically modified lectins would be assessed for toxicity, and would be unlikely to be approved.

But environmental groups argue that the approval of genetically modified products ought to be as rigorous as the drugs licensing process.

The government is currently reviewing all scientific advisory committees associated with biotechnology. E. M.

African 'flexibility' may open path to biosafety agreement

[LONDON] A United Nations conference held to reach a global protocol on the safety of genetically modified organisms opened this week in Cartagena, Colombia, with signs that African countries are preparing to compromise in order to achieve a positive outcome.

Until now, the Africa group has demanded a comprehensive legal instrument to cover the transport and impact of all genetically modified produce. "We're ready for debate. But we also want to be flexible," says Tewolde Berhan Egziabher, general manager of the Environmental Protection Agency of Ethiopia, who acts as the group's spokesman.

Egziabher says the group may be willing to compromise on its insistence that exporters would need consent from an importing country before shipping all types of genetically modified material. "We realize that there is a difference between live and dead modified organisms," he says, adding that the need for advance informed agreement may not apply to all types of modified produce.

The protocol will be finalized at the end of the special conference of the UN biodiversity convention next Tuesday (23 February). African countries are aware that failure to agree on a protocol will benefit their main opponents in the negotiations, the so-called 'Miami group' of countries, which includes the United States, Canada and other large agriculture-exporting countries.

This group, along with biotechnology and agri-chemical companies, wants a protocol restricted to regulating trade in live genetically modified organisms. They believe that a protocol that extends to genetically modified products will impede international trade.

But Egziabher says the Africa group will not shift on other contentious issues including demands for a regime for compensation in the event of transport accidents, and recognition of a country's right to refuse shipment of any genetically modified material.

Countries in the European Union, in particular, argue that countries should only be allowed to refuse shipment if genetically modified material fails an agreed scientific risk assessment of its impact, for example on biodiversity. But Egziabher says that the criteria should include the socio-economic effect of importing the material, as well as public attitudes to genetic modification.

Val Giddings, a vice-president of the Biotechnology Industry Organization in the United States, believes that few countries will agree to a treaty along these lines. He believes that the Africa group's concerns about the impact of genetic modification on biodiversity are misplaced, and claims that biotechnology benefits biodiversity.

Ehsan Masood

Scientists defend French ex-ministers in blood trial

[PARIS] The prosecution's case against three former ministers in France's 'contaminated blood' affair has been strongly challenged by a number of prominent AIDS scientists in a trial that opened in Paris last week.

Former prime minister Laurent Fabius, former minister of social affairs and national solidarity Georgina Dufoix, and her secretary of state, Edmond Hervé, are charged with involuntary homicide and involuntary harm to the physical integrity of persons. They are accused of negligence in their handling of the risk of the transmission of AIDS in the blood supply in 1985, and of having failed to take sufficient precautions to safeguard public health.

But Jean Bernard, a retired haematologist and former chairman of the French national ethics committee, told the court he did not believe that the ministers had been negligent, in particular given the uncertainties and pace of events in AIDS research at the time.

A key allegation is that the introduction of a US test manufactured by Abbott Laboratories was delayed to protect the national market for a French test, marketed by Diagnostics Pasteur. Fabius announced the decision to introduce systematic testing of blood donations for AIDS on 19 June 1985.

But Willy Rozenbaum, a pioneer of AIDS research, described a "remarkable indifference" in the scientific and medical communities in the early 1980s to the threat of AIDS. In a blow to the prosecution case, he said he had felt at the time that Fabius' decision to introduce screening was "precocious".

Rozenbaum said he was concerned that introducing screening for blood donations might actually increase the contamination of blood supplies. In the absence of a wider AIDS prevention strategy, including the creation of centres for free and anonymous testing of anyone who wanted it, he feared that introducing screening might prompt a rush to give blood from at-risk individuals as a means to get tested. And given concern that the available tests still gave false negatives, he feared that more contaminated blood would slip through the system.

Fabius denied any knowledge of the discussions of an interministerial meeting on 9 May 1985, chaired by François Gros, a former director of the Institut Pasteur, which decided that approval of the Abbott test should be held back "for some time". He claimed he had acted as quickly as possible and taken the decision to introduce systematic screening of blood while experts were still divided.

Gros has admitted that there was discussion about whether to protect the Pasteur test (see *Nature* 367, 673; 1994), but he denies

that this played a role in any delay, arguing that it was just one element in a much wider picture, the major factor having been the costs of screening.

Fabius asserted that his decision to implement screening was not based on industrial considerations. He argued, in particular, that the prosecution had incorrectly interpreted a key document. This was a note by an adviser, dated 29 April 1985, which recommended the introduction of screening, but with timing favourable to Diagnostics Pasteur.

In the margin of this note, Fabius had scribbled "I'm in favour, I will perhaps announce a decision in ten days". "It is from this that I am presented as the *chef d'orchestre* of a conspiracy favourable to Pasteur," said Fabius, arguing that his comments referred only to the need to introduce screening.

His defence also produced a letter dated 28 August 1985 from Christian Policard, the president director-general of Diagnostics Pasteur, bemoaning the fact that France was unique in having adopted a "liberal attitude to foreign tests", and alleging that all other countries had introduced protectionist measures for their own tests.

Rozenbaum, who was booed as he testified, told the court that "economic problems are permanently linked to medicine".

Luc Montagnier, a researcher at the Institut Pasteur and a member of the group that discovered the AIDS virus, said it was always possible to do things faster with hindsight, but that he felt at best only "a few weeks could have been saved" had the implementation of screening being handled differently.

A potentially lethal blow to the prosecution case came from Michel Setbon, a researcher at the Centre National de la Recherche Scientifique, who has studied screening policies in the United Kingdom, Sweden and France. Setbon told the court that "there is no relationship of cause and effect between the introduction of the diagnostic tests and the [level] of contamination".

He pointed out that the level of post-transfusion contaminations was 13 times higher in France than in the United Kingdom, even though systematic screening of donated blood was implemented in August 1985, two months earlier than in the United Kingdom.

The principal cause of the higher level of contamination in France, he said, was the failure to implement a 1983 circular intended to exclude blood donations from high-risk groups such as drug addicts. In contrast, such measures were rigorously applied in the United Kingdom and Sweden.

The trial of the former ministers is scheduled to last three weeks.

Declan Butler

Federal panel endorses Baylor fraud claim

[SAN DIEGO] One of the most legally contentious US scientific misconduct cases ended last week with the federal government deciding that a former neuroscientist from Baylor College of Medicine in Houston, Texas, fraudulently falsified data in published articles and grant applications.

A federal appeal panel ruled that Kimon J. Angelides, who was fired from the college in 1995 and subsequently moved to the University of Durham in Britain (see *Nature* 383, 107–108; 1996), engaged in “a pattern of dishonesty” in falsifying 40 figures in five articles and five applications for \$4 million in National Institutes of Health (NIH) grants.

“What the panel found here was not honest error, not disputes in interpretation of data, not preliminary results that later proved overly optimistic, not even carelessness, but rather intentional and conscious fraud,” said the panel’s report, which was released on 10 February.

Both Angelides, who quietly resigned his chair at Durham last November, and his Houston attorney declined to comment on the conclusion of a case that had been played out in numerous courtrooms. But Angelides signed an agreement with Baylor accepting the panel’s decision, under which he will be barred from receiving federal research grants for five years.

And the toll on those involved meant that, although Baylor scientists, former Angelides colleagues and attorneys threw a party to celebrate the success of their marathon legal fight, the victory was bittersweet.

“If someone had told me up front what this would take out of my life, I would have



run away,” says Susan M. Berget, a biochemist who chaired the Baylor committee that investigated Angelides’ conduct.

Universities running away from their responsibilities for investigating misconduct allegations is exactly what officials at the NIH’s Office of Research Integrity (ORI) fear may happen in the aftermath of the closely watched Angelides case.

When Baylor fired Angelides following a three-year investigation that found he falsified research, Angelides sued Baylor, its investigating committee members and several former students and postdocs who gave evidence against him. Angelides alleged that the accusers had slandered him and destroyed his career.

After a series of legal manoeuvres, the fact that the trial on Angelides’ allegations began last month in the state court was of particular

concern to Baylor and federal officials. Baylor and the federal agencies believe the university and its investigating scientists should be immune from such lawsuits, as the college was fulfilling the NIH’s requirements for grant monitoring.

But just as Angelides was completing the presentation of his case in the Houston courtroom, the final federal administrative decision was issued. A research integrity adjudication panel of the Departmental Appeals Board for the NIH issued a 172-page report confirming the findings of Baylor and ORI.

With the report having probably destroyed his chances in the civil lawsuit, Angelides agreed to dismiss the lawsuit and entered into a settlement with Baylor. As part of the settlement, Baylor agreed to pay \$500,000 of Angelides’ legal bills.

Baylor feared that Angelides and his attorney, James V. Pianelli, would try to fight the panel’s decision in the federal court. Angelides’ legal costs were reportedly around \$800,000; Baylor’s were more than double that amount.

“I hope this case serves as a wake-up call to legislators and regulators for the need to enact laws protecting institutions and their investigative committees from lawsuits of this nature,” says Baylor’s attorney, Gerard G. Pecht.

To prevent accused scientists from using the courts to block or undermine investigations of scientific integrity, Pecht argued that academic institutions and their investigating committees should have limited immunity from such lawsuits and that any legal complaint be held in abeyance until the federal administrative review process is complete.

NIH officials argued similarly during earlier legal proceedings in the case. ORI officials say they hope to secure legislation providing the necessary safeguards.

Chris B. Pascal, acting director of ORI, said Angelides’ lawsuit “was a threat to the ability of institutions to conduct these necessary investigations. Hopefully, this decision will discourage other lawsuits of this type.”

Pascal also noted that there had been threats to those who testified for Baylor. “Scientists involved in the investigation were forced to take valuable time out from their research to defend themselves. Some scientists were falsely accused of wrongdoing, forced to live under a cloud of suspicion. This decision vindicates them.”

Berget says that discussions with jurors after the settlement left no doubt in her mind about the need for institutional diligence in addressing misconduct allegations. “Jurors easily bought Pianelli’s argument that scientists do this (falsify data) all the time,” says Berget. “That tells us we have to conduct thorough investigations.”

Rex Dalton

Anticancer activity of endostatin redeemed

[WASHINGTON] The National Cancer Institute (NCI) announced last week that it had finally succeeded in replicating the high-profile experiment of Harvard researcher Judah Folkman with a compound that shrinks tumours in mice by choking off their blood supply.

The NCI’s success with mouse endostatin – just months after it had announced publicly that it could not replicate the results – led the institute last week to begin soliciting proposals for phase I (safety) trials of the human version of the drug.

NCI officials said that the reason they succeeded where they had failed earlier

was that they had visited Folkman’s laboratory at the Children’s Hospital in Boston, Massachusetts, where they learned his techniques for dealing with the volatile natural compound.

The announcement more than doubled the share price of EntreMed, Inc., of Rockville, Maryland. The small company, which has obtained from the Children’s Hospital in Boston the rights to develop the drug commercially, had seen its shares tumble to \$12.77 earlier last week after Bristol-Myers Squibb announced that it was withdrawing its support for EntreMed’s development of a

similar compound, angiostatin.

They said it was not clear that angiostatin could be made in large enough amounts for human testing. The NCI announcement sent the share price up to \$25.71.

The potential of both drugs – which in Folkman’s experiments have worked most powerfully in combination – came into sharp public prominence last May after the *New York Times* ran a front-page story quoting Nobel laureate James Watson as saying that Folkman would “cure cancer in two years” (see *Nature* 393, 104; 1998). Meredith Wadman

Higher status for 'research risks' office?

[WASHINGTON] The future of the office that protects human subjects in much US biomedical research is likely to be decided by a report soon to be delivered to Harold Varmus, the director of the National Institutes of Health (NIH).

Whether the Office for Protection from Research Risks (OPRR) should remain in the NIH, or be elevated to the office of the secretary of the Department of Health and Human Services (DHHS), is being considered by a six-member panel put together by Varmus last autumn.

The two co-chairs of the panel declined to comment on the report's contents, pointing out that these are not final. But there is belief in Washington that the group is likely to recommend the office's elevation to the DHHS.

The group is also charged with deciding whether the office needs more authority to

ensure that thousands of institutions and government workers conducting DHHS-funded research worldwide comply with regulations protecting animal and human subjects.

The panel was planning to debate its draft report in a conference call scheduled for yesterday (17 February). It will then go "as soon as humanly possible" to Varmus, says Nancy Neveloff Dubler, one of the panel's co-chairs, a professor of bioethics at the Albert Einstein College of Medicine in New York.

According to Dubler, although the issue of research with human subjects has attracted relatively little attention for a long time, it is now under increasing scrutiny. "This comes at an important moment," she says.

For instance, the ethics of psychiatric research in which subjects receive drugs that induce symptoms, or are withdrawn from

medications and become symptomatic, have received widespread media attention. As a result, the National Institute of Mental Health has set up a panel to provide an additional layer of scrutiny for such experiments.

The OPRR's effectiveness was questioned at a Congressional hearing last year (*Nature* **393**, 610; 1998), where Congressman Christopher Shays (Republican, Connecticut) called it "pathetic" that the office had only one full-time investigator into human subjects on its staff of 29 (three employees also work as investigators part-time), the rest primarily being administrators.

The weaknesses of the office were also documented in a separate report last year by the DHHS inspector-general, and a report two years earlier by the General Accounting Office (GAO). Among other problems, critics point out that the office's budget, at \$2.1 million in 1998, has declined as the amount of research it is required to monitor has grown.

But perhaps the chief concern of critics, including the GAO, is what they call an inherent conflict of interest in the office operating within the NIH, where it is in the Office of Extramural Research. "It is very difficult to monitor human subjects' protection from within an agency that has as its primary mission the advancement of research," says Arthur Caplan, director of the Center for Bioethics at the University of Pennsylvania.

The GAO report, for instance, stated that in one case it took five years for intramural scientists at the NIH to implement corrective actions that had been required by the OPRR.

At a Congressional hearing in 1997, Varmus denied to Shays that a conflict existed, arguing that the OPRR had no vested interest in research progress. That Varmus set up the committee — which will officially report to his advisory committee — appears to signal his readiness for the office to be moved.

Some argue that placing the OPRR directly under Donna Shalala, the secretary of health and human services, would still not guarantee the office the independence it needs in the face of the rapid expansion of US biomedical research.

John Fletcher, for example, emeritus professor of bioethics at the University of Virginia, has proposed that the office be an independent entity within the government, like the Nuclear Regulatory Commission. "This is a national issue," he says. "The NIH is a major research institution, but it is one among many. The agency would have more respect and influence if it were independent."

Others argue that, wherever it is located, the OPRR needs a significantly larger budget. "A 'mom and pop' store is no longer adequate when a supermarket is required," says Caplan.

Meredith Wadman

Congressman aims to punish cloning research

[WASHINGTON] In a warning shot in the battle on human cloning, a conservative Republican has introduced a bill banning federal payments to any business, institution or organization that "engages in human cloning or human cloning techniques".

Congressman Ron Paul (Republican, Texas), a former obstetrician and gynaecologist, says his bill has been crafted to avoid inhibiting research: "It just means that universities and medical centres that set out to clone people get their funds cut off." But biomedical researchers have described the proposal as "massively punishing".

The bill, which was introduced two weeks ago, has so far won no co-sponsors, and Paul, a junior member of Congress, has little power to move it without patronage from more senior Republicans. However, it is symbolic of conservative opinion on cloning.

David Korn, senior vice-president for biomedical and health sciences research at the Association of American Medical Colleges, says the bill goes further than the existing ban on human embryo research, as it would



Paul: wants funding ban.

cut all federal funding to institutions where researchers use somatic-cell nuclear transfer, even if they do so with private funds. (The present ban allows privately funded embryo research.)

Furthermore, says Larry Goldstein, a professor of cellular and molecular medicine at the University of California, San Diego, the definition of 'human cloning' in the bill is so ambiguous as to potentially prohibit stem-cell research.

The bill defines human cloning as "making an identical ... copy of the genetic material of an individual ... so as to cultivate one or more new human cells which could, if not otherwise engineered, develop into a new individual human being".

While it is believed that stem cells isolated from

unused embryos left over from fertility treatments could not grow into a human if they were implanted in a uterus (*Nature* **396**, 104; 1998), this is not known for certain and cannot be proved without experiments that are themselves considered unethical. As a result, says Goldstein, the bill would at the least allow legal challenges to stem-cell work.

Paul disputes that interpretation, saying that stem-cell research would be protected. But he makes no apologies for seeking to ban any cloning that would produce a human embryo not destined for life, even though this technique could aid the development of cell and tissue therapies.

Meanwhile, the American Society for Cell Biology met around 30 representatives of scientific and patient groups last week to discuss their approach to the stem-cell research issue on Capitol Hill. Forty-five groups have signed a letter to Congress applauding a recent legal opinion by the Department of Health and Human Services that federal funding of stem-cell research is allowed under the current law (*Nature* **397**, 185; 1999).

M. W.

Early American goes for tests — at last

[WASHINGTON] A set of 9,000-year-old human remains found near Kennewick, Washington State, nearly three years ago will finally undergo a detailed scientific examination next week to try to determine its origin.

The tests follow a protracted legal battle involving archaeologists, Native American tribes and the US government. Their outcome could have far-reaching implications for scientists studying the early settlement of America.

Two college students discovered the skull of 'Kennewick Man' in a shallow stretch of the Columbia River in 1996, and archaeologists retrieved 300 additional bone fragments from the site. Local Native American groups claimed the remains under the 1990 North American Graves Protection and Repatriation Act (NAGPRA), with the intention of reburial of them.

But a group of eight scientists sued the US government to halt the reburial, arguing that they had a right to study the remains under the first amendment to the constitution, which protects freedom of expression.

So began a controversy described as "sur-realist" by Robson Bonnicksen, an Oregon State University archaeologist and one of the plaintiffs. The Asatru Folk Assembly, a group that practises ancient Nordic religion, also claimed rights to Kennewick Man on the grounds that his apparently Caucasian features suggest a European origin.

The bones have been moved at least twice, and are now at the Burke Museum of Natural History and Culture in Seattle, where next week's tests will be conducted. The examination will be done by five government-appointed scientists, including two physical anthropologists, two soils experts, and a specialist in arrowheads and other lithic objects.

Francis McManamon, chief archaeologist for the Department of the Interior, which administers NAGPRA, says the first investigations will include physical measurement and analysis, but no DNA sampling, radiocarbon dating, or other destructive techniques opposed by the tribes.

A US district court has ordered the government to determine whether Kennewick Man was Native American and to which modern tribe he might be related. A radiocarbon date taken for one of the fragments yielded a date between 9,200 and 9,600 years ago.

Bonnicksen says that, although he and his fellow archaeologists support NAGPRA's goal of giving Native Americans control over their cultural artefacts and ancestral remains, the law is out of step with modern thinking. Scientists now agree that there was more than one migration of people from Asia to America.

The biological descendants of some of these 'palaeoamericans' have long since died



Skull and cross words:
Kennewick Man has been the centre of controversy.

out, he says, and modern tribes may therefore bear no relation. Many Native American tribes, on the other hand, claim to have lived on their ancestral lands for ever.

Underlying the Kennewick Man debate, Bonnicksen believes, is an "epistemological

conflict" between Western science and Native American creationism. "We're not going to resolve those differences," he says. But he accuses the government of unfairly siding with the Native Americans.

The interior department takes the position that any artefacts or remains that predate the arrival of European explorers qualify as 'Native American' under NAGPRA. "Scientists find that definition totally unacceptable," says Bonnicksen, and "based solely in politics".

McManamon agrees that the law did not anticipate having to account for palaeoamericans unrelated to modern tribes. A bill introduced in the last Congress

by Washington Republican Doc Hastings would have revised the act to allow scientific study of remains "for which a cultural affiliation is not readily ascertainable".

Until such a bill passes, however, the interior department remains stuck in the middle. Some Native Americans are upset that any tests, even non-destructive ones, will be conducted next week. And scientists complain that they will not be enough.

"I don't see how this non-destructive stuff is going to answer the question," says Bonnicksen. He also worries that the results will be censored, and that there will be no outside peer review. Observers from both sides will be allowed to monitor the tests.

McManamon hopes that analysis of soil adhering to the bones will tie them to a particular geological setting that can help to date them. If the non-destructive tests cannot determine the age of the fragments or their connection to modern tribes, the interior department plans to move on to DNA and radiocarbon tests, after consulting the tribes.

Despite his complaints, Bonnicksen says the decision to do any tests is "a huge step forward". He thinks the controversy has resulted in an "enormous education process," with many more people now understanding that the settling of America was more complicated than previously thought. Fewer than 15 skeletons of Kennewick Man's age have been found in America.

Tony Reichhardt

Union backs Toronto researcher's claims

[MONTREAL] A research associate who claims that racial discrimination by the University of Toronto led to his failure to win a tenure-track position was unfairly treated by the university, concludes a report by the Academic Freedom and Tenure Committee of the Canadian Association of University Teachers (CAUT).

But the university has criticized the report as "deeply flawed", and argues that its recommendations "are either unnecessary or offensive to academic values". It says its own investigation into Kin-Yip Chun's case found no evidence of racial discrimination, and that, although it did find "irregularities and ambiguities" in his original appointment, it has "offered appropriate remedy" (see *Nature* 392, 638; 1998).

The CAUT report recommends that the university offer Chun "an opportunity to return to work at the university with genuine job security on a continuing basis at a rank and salary commensurate with his actual experience".

It says that he should also be given a fair chance to compete for any position in seismology. If a negotiated settlement is not

possible, the parties should submit the matter to binding arbitration, which would include the possibility of awarding Chun a tenure-track position.

Chun, a seismologist, was passed over four times in applications for tenure-track positions at the university, and in 1994 was escorted off the campus by a plain-clothes policeman and banned from the campus. The university says it has offered Chun a settlement "consistent with the CAUT recommendations" and "continues to propose a mediated settlement. Chun has persistently refused all offers."

Although the CAUT report "found no evidence of direct discrimination", it said it received "evidence of systemic discrimination on grounds of race or ethnic origin". This included the "ethnic and gender composition" of the department to which Chun had applied for posts, which was "overwhelmingly white and male".

The university rejects these "speculations". It says it "will not award a tenure stream position as the result of arbitration, but only by peer adjudication through normal processes". David Spurgeon

US panel says research can be assessed

[WASHINGTON] A committee of the US National Academy of Sciences said last week that an effective assessment of research performance must take into account three factors: the quality of the research, its contribution to world leadership in various fields, and its relevance to the goals of funding agencies.

The Committee on Science, Engineering and Public Policy (COSEPUP) said that, despite serious doubts, research performance can be assessed — applied research easily, basic research with more difficulty.

The report is a response to attempts by research agencies to implement the Government Performance and Results Act (GPRA), which requires federal agencies to measure how their programmes benefit taxpayers.

The committee, chaired by Phillip A. Griffiths, director of the Institute for Advanced Study in Princeton, cites the pros and cons of common measures of research quality such as bibliometric analysis, economic rate of return, peer review, case studies, retrospective analysis, and benchmarking.

All feature in evaluations. But a magic element is needed, it says: “expert review”, an

assessment of the “quality–relevance–leadership troika”. Making such an assessment requires unique and seldom-used combinations of expertise, it adds.

The 80-page report stresses the development of research talent in fields critical to agency missions. Most performance plans have lacked such goals, says COSEPUP. But it emphasizes that education and training are essential in any research performance plan.

COSEPUP warns that focusing on short-term results can undermine the whole purpose of evaluation. This echoes the National Science Foundation's struggles to comply with GPRA. The foundation was the first agency to release a performance plan geared to its fiscal year 2000 budget proposals.

The committee also suggests extending GPRA by developing assessments of programmes involving several agencies — such as global change research, or the information technology initiative announced earlier this month in the Clinton administration's 2000 research and development budget.

COSEPUP says the administration should designate lead agencies to ensure that

interagency programmes are well coordinated, a task that usually falls to the Office of Science and Technology Policy (OSTP) as the operating arm of the National Science and Technology Council.

OSTP has submitted its own GPRA strategic plan, but has yet to release any performance plan. It may need to if GPRA assessors in Congress adopt COSEPUP's coordination recommendation. In a final recommendation, COSEPUP says that researchers must take a leading role for the whole process to work.

In its original form, agencies had to submit strategic plans to GPRA by September 1997, performance plans by spring 1998, and performance reports by March 2000. All performance plans are now in, although some remain in outline form, and many are being revised.

Committee member Morris Tanenbaum, former chairman of the AT&T telecommunications corporation, says GPRA is critical to give researchers a better understanding of government and Congress a better understanding of research.

Wil Lepkowski

Giant radio telescope will boost search for extraterrestrial life

[BOSTON] Scientists at the University of California at Berkeley and the SETI Institute — named after the Search for Extraterrestrial Intelligence — plan to build a new radio telescope that is described as a ‘paradigm shift’ in the design and construction of such instruments.

The proposed ‘One Hectare Telescope’ (1hT), consists of an array of 500 to 1,000 small satellite dishes with a combined collecting area of 10,000 m² (one hectare). This would make it the world's largest telescope devoted mainly to SETI. The 1hT is intended to be a prototype for the Square Kilometer Array (SKA), which is being planned by an international consortium of radio astronomers for general research.

The 1hT effort was launched last summer after agreement was reached between Berkeley and the SETI Institute. “The germ of this project goes back to late spring when we were thinking about how to build SKA and decided that an array of small dishes would be the way to go,” says project scientist John Dreher, an astronomer now based at both institutions. “But the Square Kilometer Array may not be built until 2010, and we didn't want to wait that long.”

Preliminary funding has been secured through 1999 and part of 2000, and some radio dishes have already been purchased. Researchers soon hope to experiment with an array of ten dishes, probably installed at Berkeley's Hat Creek Observatory.



Big ears: the one-hectare telescope will be a prototype for a square-kilometre array.

The one-hectare array could be built by 2004, either at Hat Creek or at other ‘radio-quiet’ sites in California currently under consideration, assuming that the SETI Institute can raise the necessary funds.

By using commercial technology, the team hopes to keep the costs of the array below \$25 million, a tenth of the cost of a conventional telescope of comparable size.

In contrast to a big telescope, a small dish can look at a big chunk of the sky. “That chunk might contain ten stars plus other astronomical sources, and we can observe all of them simultaneously,” says SETI Institute astronomer Jill Tarter. “So my [SETI] search goes ten times more quickly, or I can spend ten times as long on each target.”

With the new telescope's enhanced sensitivity, Tarter and her colleagues can expand their “target list” from 1,000 Sun-

like stars to 10,000 or even 100,000. The 1hT's large field of view will allow radio astronomers to study large-scale sources. “It's big enough to see the whole Sun,” says Dreher. The instrument will also be well suited to looking at pulsars, comets and a variety of dim sources, such as gas clouds.

The telescope will reveal the “low-frequency Universe” at greater distances than before, says James Moran, a radio astronomer at the Harvard-Smithsonian Center for Astrophysics. “But its greatest contribution will be the development of really cheap technology that could be used in the Square Kilometer Array.”

The SKA, in turn, will “open doors for us”, says Leo Blitz, director of Berkeley's radio astronomy lab. “It will enable us to measure the primordial hydrogen out of which the first galaxies formed.”

SETI investigators could conduct deeper surveys using the SKA, bringing more stars into range, and it would be sensitive enough to pick up ‘leakage’ radiation from stars — meaning that researchers would no longer be limited to searching for civilizations that are intentionally beaming our Solar System.

“That will bring the SETI search to the realm where we can actually determine the frequency of civilizations like ours,” Blitz says. “Of course, most people engaged in this search are not after statistics. They hope to find something that shows we are not alone.”

Steve Nadis

Mexico to overhaul R&D funding system

[MEXICO CITY] The administration of Mexican president Ernesto Zedillo is seeking a wide-ranging overhaul of the way public funding for research is allocated and administered.

A proposed law due to be considered by the Mexican Congress this spring is intended to promote and rationalize the support of scientific and technological research, in particular by setting up a unified, government-wide mechanism for planning and budgeting for research and development (R&D).

The consolidated R&D budget will comprise the planned expenditures in each of the relevant ministries. The aim is to allow the agencies that support R&D to interact, and to commit individual ministries to a specified level of support.

Pablo Rudomín, Zedillo's science adviser, says this will prevent funding for science from being edged out by ministries facing more pressing short-term funding demands. At present, there is nothing to prevent a ministry from re-allocating money originally designated for research and development.

The proposed legislation results from a joint initiative by the leaders of the president's council of science advisers, the Mexican Academy of Sciences, and CONACYT, the country's main research funding agency, at Zedillo's request.

According to Rudomín, Zedillo accepts that science needs more money, but thinks it must also "change structures, remove barriers and optimize the use of resources".

Although it is not mentioned in the proposed legislation, Zedillo plans to create and preside over a ministerial-level 'science and technology cabinet'. It is hoped that this cabinet and the consolidated programme will together allow national objectives for the public support of R&D to be defined, and ultimately lead to growth in the R&D budget.

Science and technology will be given a statutory voice in government through a 'permanent forum' of scientists, academics and industrialists, who will advise both the executive and legislative branches of government on policies for supporting R&D.

The new law will also change the way public research centres are administered. The laws controlling these centres were originally designed for state-owned industries, such as the oil company PEMEX, and are widely considered too restrictive for organizations carrying out scientific research.

The Centre for Research and Advanced Studies of the National Polytechnic Institute, for example, which is funded by the Ministry of Education, is audited several times a year, and its managers are not permitted to move money from one budget line to another, for example, to buy computers instead of chemicals.

The new law would grant these centres



Zedillo: plans to set up a 'science cabinet'.

In addition, the ministries operating such centres would be given increased flexibility in allocating their research funds. Unlike CONACYT, which gives research grants to scientists in any institution, the ministries have traditionally funded research only in their own centres. The new law would encourage them to invite and fund proposals from anywhere, including the universities.

The legislation is the most visible product so far of a collaboration agreement signed two-and-a-half years ago by the president's council of science advisers, the Mexican Academy of Sciences and CONACYT. Previous accomplishments of this alliance have included a presidential decree exempting research materials from import tax, and a national registry of scientists and engineers.

Members of the alliance consulted widely with leaders of the scientific community before drafting the legislation, to develop a consensus for the proposed changes. However, many Mexican scientists are unaware of

administrative autonomy, equivalent to that of universities. It would also promote links between public research centres and private industry by removing the current cap on each centre's budget — thereby providing an incentive for the centres to raise funds from private sources.

the law or uncertain of its detailed provisions. Oscar Prospéro-García, a neuroscientist at the National Autonomous University (UNAM), says that, in principle, it seems to be a good law, but "we don't yet know how it will work".

Other scientists would have liked a commitment in the law to a specified level of support for research and development. "What they should have done is commit to a percentage of gross domestic product [GDP]," says Mariano López de Haro, a physicist at UNAM's Energy Research Centre who is a member of the academy of sciences.

The government spends less than 0.3 per cent of Mexico's GDP on research and development. This figure has not changed during the Zedillo administration, despite early promises to increase it.

Rudomín says that consideration was given to including such a commitment, but the lawyers who drafted the legislation counselled against it on legal grounds. "We would all love — and we need — to see a substantial increase in support for research," he says. "But for this we need science to become part of our culture, so that people will recognize the importance of increased support."

The bill's proponents hope that Mexico's upper house (the Senate) will consider the legislation shortly after it re-convenes in March. Rudomín says he feels Congress will look favourably on the bill, at least in its broad outlines, and that it won't be treated as a partisan issue. "We have insisted that this goes above parties — it's something that the country needs."

Laura Garwin

Switzerland adopts transplant guidelines

[ZURICH] New guidelines on procedures for organ transplantation are being incorporated into the Swiss constitution following their approval in a national referendum on 7 February.

The guidelines will establish general principles for a planned federal law regulating transplant medicine. At present, some Swiss cantons have legislation on transplants, while others have none at all.

The move prohibits organ donation for money, and requires fair and equal access to donated organs, regardless of sex, race or wealth. But it fails to take a firm stance on xenotransplantation, the definition of death, or transplants of embryo cells and tissues.

Environmentalists, animal activists and opponents of genetic engineering had campaigned against the constitutional proposal, mainly because it does not include an outright ban on xenotransplants. That issue will be regulated by a new federal law on transplant medicine being developed by

the federal office of health. Thomas Zeltner, its director, says he does not expect the law to come into effect before 2002.

According to sources in Bern, a ban on xenotransplants is highly likely. The parliamentary health committee has voted in favour of a moratorium, and the government wants to outlaw the use of animal organs for transplants in general.

But clinical trials of xenotransplants offering potentially important therapeutic benefits could be allowed after special permission. Swisstransplant, a charity that currently regulates all transplant activities in Switzerland, welcomes the move as an urgent step towards clear and unified rules throughout the country.

But some experts warn that a ban on paid organ donation will not be helpful in resolving the chronic shortage of organs in Switzerland. Last year, 920 patients needing a transplant remained on waiting lists, while 28 died untreated.

Ulrich Bahnson

Japan aims to launch 1,000 biotech companies in 10 years

[TOKYO] Japan plans to expand its biotechnology market by a factor of 25 by the year 2010, under a programme to be launched by the country's science-related ministries and agencies. According to a government announcement last week, the plan will support more than 1,000 new biotechnology-related companies within the next decade.

Five ministries and agencies will be involved, including the Science and Technology Agency, Ministry of Agriculture, Forestry and Fisheries, and the Ministry of International Trade and Industry.

The programme is expected to create a market worth ¥25 trillion (US\$212 billion). It will promote the commercial application of research into human, animal and plant genomes by creating vast databases containing such genetic information.

Italian institute to link astronomy observatories

[MUNICH] The Italian government has approved the creation of a National Institute for Astrophysics (INAF), which will merge

the administration and coordinate the research activities of Italy's 15 previously independent astronomy observatories.

Modelled on INFN, the highly regarded national institute for particle physics, INAF will distribute the budgets of each observatory, develop strategic research plans, and represent the interests of Italian astrophysicists at international organizations such as the European Southern Observatory.

Cadbury to chair UK's Wellcome Trust

[LONDON] Sir Dominic Cadbury, chairman of the chocolate and soft drinks company Cadbury Schweppes, is to be the new chairman of the medical charity the Wellcome Trust. He will succeed Sir Roger Gibbs, whose decade at the trust's helm has seen its investment portfolio grow by almost 20 per cent each year, exceeding that of most market indices.

The trust's assets of £11.7 billion (US\$19 billion) are said to make it the largest grant-giving charity in the world. Cadbury will become a governor of the trust on 1 June, before taking over as chairman on 1 January 2000. He is a past chairman of the education and training affairs committee of the Confederation of British Industry.

US launches trial of AIDS vaccine in Uganda

[WASHINGTON] The US National Institute of Allergy and Infectious Diseases (NIAID) announced last week that it has launched the first clinical trial of an AIDS vaccine in Africa. The phase one (safety) trial of a canarypox-based HIV vaccine, made by Pasteur Mérieux Connaught, will enrol 40 uninfected volunteers at the Joint Clinical Research Center in Kampala, Uganda and the Uganda Virus Research Institute in Entebbe.

Separately last week, VaxGen, Inc. of San Francisco, California announced that Thailand's Ministry of Public Health has approved the launch of a phase three (large scale efficacy) trial of the company's gp120 vaccine — the first efficacy trial outside North America. The Bangkok trial will enrol 2,500 uninfected, high-risk volunteers.

Canada to boost Earth sciences collaboration

[MONTREAL] A Canadian federal programme to support advanced Earth sciences research and training has been announced by Ralph Goodale, minister of natural resources, and Ronald Duhamel, secretary of state for science, research and development.

Collaboration will be sought among

scientists and engineers in industry, the federal government and universities. The Earth sciences sector of the natural resources department and the Natural Sciences and Engineering Research Council will each invest Can\$500,000 (US\$335,000) in the programme. The private sector is expected to contribute a similar amount. All grant applications must involve at least one private-sector or industrial partner.

NZ bill aims to give apes same rights as humans

[TOKYO] New Zealand's parliament is about to vote on a bill conferring the equivalent of human rights on great apes such as chimpanzees, gorillas and orang-utans, giving the animals protection from all but the most benign experiments.

If the bill is passed, the Great Ape Project — a campaign led by 38 of the country's scientists, lawyers and philosophers — will make a call for the rights of great apes to be included in a United Nations declaration.

The proposed bill has been welcomed by animal rights groups. But some biomedical researchers, concerned that such a move would eventually target all research animals, say that inhumane treatment of apes can be prevented without addressing the notion of animal rights.

Mbeki makes plea for 'African renaissance'

[CAPE TOWN] South Africa's deputy president, Thabo Mbeki, last week appealed to delegates at an international conference in Cape Town on women, science and technology for sustainable human development to contribute to the "African renaissance". The meeting coincided with the second general assembly of the Third World Organization for Women in Science (TWOWS).

A statement agreed by the conference urged "organizations, institutions and governments around the world to take every possible measure aimed at achieving a full involvement of women in science and technology for the benefit of our people and our societies". The president of TWOWS, Lydia Makhubu, vice-chancellor of the University of Swaziland, said the conference had demonstrated widespread awareness of brightening prospects for postgraduate training for Third World women.

\$50m data network will link Europe's researchers

[LONDON] A US\$50 million contract to set up an advanced, high-speed fibre optic network linking European research centres has been

won by KPNQwest, a company owned by Qwest Communications in the United States and KPN of the Netherlands.

The first phase will link seven European cities in France, Germany, Holland and Belgium. Five additional rings connecting up to 40 cities are expected to be built by 2000. The network will be able to transmit up to two terabits of data per second, and will eventually connect to a similar fibre optic network in the United States.

Solar power advances win Australia prize

[SYDNEY] This year's Australia Prize in Energy, Science and Technology has been awarded to two researchers for their pioneering work on increasing the efficiency of harnessing solar energy. The prize, worth A\$350,000 (US\$225,000), was presented this week by the Australian prime minister, John Howard.

Martin Green and Stuart Wenham used thin films of multi-layered cells of silicon on sheets of cheap glass to achieve a record 24.5 per cent efficiency and greatly reduced production costs. Green began his quest in 1974 at the University of New South Wales in Sydney, where Wenham has worked with him since 1983. Wenham predicts that the price of solar power will soon drop from A\$4 (US\$2.60) per watt to A\$1 per watt.

Don't count on World Bank initiatives

Sir — As a Chilean scientist working overseas, I read with interest your articles about the initiatives of the World Bank to develop science and technology in Third World countries^{1,2}. These initiatives include lending funds to create elite centres called Millennium Institutes. Prototype institutes will be set up in Chile. I believe this is the wrong approach.

I collaborate with scientists in Third World countries and am familiar with the vagaries of the development of science in Latin America. I wish to put forward the dissenting notion that publicly funded policies to develop science and technology in Third World countries should not differ from those proven successful in industrialized countries. These include: a focus on funding innovative investigator-originated peer-reviewed research; transparency of all procedures and full public accountability of those involved in the funding process; and societal inputs to create economic and social conditions that ensure the availability of resources for the development of science and technology.

Third World countries need to narrow the knowledge gap that separates them from industrialized countries. But the creation of one or two elite Millennium Institutes in each country will do little to close that gap. In most Third World countries the technological and scientific gap has widened during the past 20 years, in part as a result of the economic and social engineering prescriptions of the World Bank and the International Monetary Fund (IMF). These prescriptions in general stimulate the creation of export economies based on unskilled labour, intensive exploitation of natural resources, and export of raw materials without industrial transformation and technological

innovation. These prescriptions undercut the creation of market conditions basic to scientific and technological development³.

In Chile, for example, recent economic growth has been based mainly on exporting minerals, fish and fish meal, timber and fruit, all products with little or no industrial transformations, that do not require technological innovation, and whose production has generated environmental degradation. The World Bank and the IMF foster policies that, under the guise of 'restructuring and modernization', shift government funds from activities essential for the development of science and technology, such as education and public health, towards servicing foreign debt⁴. They further weaken a poor country's capacity to develop the scientific and technological bases needed for economic development.

In Chile, these policies have led to deterioration of the educational system and crumbling infrastructure. The increases in tuition fees that have resulted from World Bank and IMF-supported policies have excluded many talented individuals from the system. The same policies have caused teachers and scientists to live on the margins of poverty. The implementation of these economic policies in Chile was also accompanied by the destruction of democratic institutions and the abolition of human rights. The incomplete recovery of these institutions and human rights still scars and undermines scientific work⁵.

In this context the creation of Millennium Institutes in Chile, and perhaps in other countries, appears to be more a cosmetic than a well designed initiative. This plan fails to address the economic and social foundations of the host of problems that hamper the development of education and science in Third World countries.

Another matter of concern is that the directors of these institutes will be selected by scientists from outside the country. In the case of Chile these institutes will be funded with loans that will eventually be paid by the Chilean treasury. It is therefore worrisome that the manner of selection will remove accountability from the Chilean scientific community and Chile's democratic processes. Despite the undermining of education, science and technology resulting from the policies of the Pinochet dictatorship, and from the relentless 'restructuring and modernization' induced by World Bank and IMF economic and social engineering, the Chilean scientific community within the country and in the diaspora has reached a stage that makes outside tutelage unnecessary.

Who is in a better position than Chilean scientists and politicians to determine whether the focus of the research is "of direct relevance to that country's needs"? As you note, "the Chilean government will not decide the disciplines in which the institutes will specialize", making it all the more necessary that the Chilean scientific community should make these decisions.

The closing of the knowledge gap in Third World countries will require drastic modifications of present economic policies. The consensus to create these changes will require the participation of scientists and politicians in each country.

Felipe Cabello

*New York Medical College,
Valhalla, New York 10595, USA*

1. Macilwain, C. *Nature* **396**, 711 (1998).

2. Masood, E. *Nature* **397**, 6–7 (1999).

3. Danaher, K. *Fifty Years is Enough: The Case Against the World Bank and the IMF* 51–61 (South End Press, MA, 1994).

4. Lurie, P., Hintzen, P. & Lowe, R. A. *AIDS* **9**, 539–546 (1995).

5. Collins, J. & Lear, J. *Chile's Free-market Miracle: A Second Look* 93–148 (Inst. for Food and Dev. Policy, Oakland, CA, 1995).

Harvest time is over for UK agricultural research

Sir — English agricultural researchers can be pleased with the international comparison of research impact (*Nature* **396**, 615–618; 1998). Looking at 47 subjects from 1988 to 1996, the analysis shows that agriculture was one of only five subjects in which England came top of the list. Within England, agriculture came second only to pharmacology/pharmacy.

However, towards the end of that period and subsequently, there has been decreasing emphasis on agricultural research in England. The former Agricultural and Food

Research Council has omitted the word 'Agricultural' from its name and changed its orientation accordingly. Some university departments of agriculture have become subsumed into biological schools and some are unable to fill vacant agriculture chairs. The Ministry of Agriculture, Fisheries and Food (MAFF) has terminated its Postgraduate Agricultural Studentship Scheme for training agricultural researchers.

The former MAFF extension service has been privatized: will its researchers now have much interest in diverting time from income generation to publication in refereed journals?

In the light of the above changes, it is difficult to see the international success of

English agricultural research being maintained.

Peter S. Kettlewell

*Crop and Environment Research Centre,
Harper Adams University College,
Newport, Shropshire TF10 8NB, UK*

Evolutionary forces behind human infertility

Sir — Westendorp and Kirkwood¹ studied data on the British aristocracy and conclude (probably correctly) that there is an evolutionary trade-off between longevity and fecundity. There is another significant

conclusion to be drawn, and that is that humans have a strong hereditary predisposition to infertility.

The tendency for low fecundity among the English aristocracy was in fact first studied systematically by Galton in 1869 (ref. 2). There was general concern at the time at the rate of extinction of English hereditary peerages. Galton examined the links between social status and reduced fecundity, and concluded that it was largely due to the tendency for peers — and the sons of peers — to marry heiresses as a means of accruing estates. He speculated that in a patrilineal society heiresses are more likely to arise in small families of reduced fertility, and that marriages with such women would also have a tendency for low fecundity.

Fisher³ later analysed Galton's data and other genealogical findings in some detail, and concluded that human reproductive success is extremely unevenly distributed and therefore subject to very strong selective pressures. In the 1912 Australian census, for example, 50 per cent of the children were the offspring of one in nine of the men and one in seven women. Three-fifths of all children that were born died unmarried and 11 per cent of marriages were sterile⁴.

While Fisher and Galton's writings were tarnished because of their links with the eugenics movement, one clear message is that subfertility is endemic within human populations — albeit hidden in ancestral communities by child-sharing and other devices such as serial polygamy⁵. Moreover, reduction of investment in reproduction is a powerful force for wealth consolidation within a family: a concept that Fisher³ traced back to Hesiod in the eighth century BC! The link between longevity and reduced fecundity is entirely consistent with the disposable-soma hypothesis discussed by Westendorp and Kirkwood¹.

In an era when infertility is increasingly emergent as a social problem and is now eminently treatable by technological means, we should perhaps be aware of the evolutionary forces that may have helped amplify it.

For males, the strong selective pressure for critical genes controlling fertility on the Y chromosome⁶, coupled with highly uneven reproductive success between individuals in any generation, should be considered in any attempt to reconstruct genealogies based on Y sequences. Rather than differential migration rates between women and men, this could possibly explain part of the discordant convergent times for human Y sequences and mitochondrial DNA⁷. While Y chromosome variations are generally considered to be neutral, close scrutiny of actual patrilineal reveals that long-term

male reproductive success is strongly influenced by politics and social dominance⁸.

Jim Cummins

*Division of Veterinary and Biomedical Sciences,
Murdoch University, Western Australia 6150,
Australia*

1. Westendorp, R. G. J. & Kirkwood, T. B. L. *Nature* **396**, 743–746 (1998).
2. Galton, F. *Hereditary Genius: An Inquiry into its Laws and Consequences* (Macmillan, London, 1869).
3. Fisher, R. A. *The Genetical Theory of Natural Selection* (Dover Press, New York, 1958).
4. Heron, D. *Biometrika* **X**, 419–420 (1914).
5. Neel, J. V. & Weiss, K. M. *Am. J. Phys. Anthropol.* **42**, 25–51 (1975).
6. Lahn, B. T. & Page, D. C. *Science* **278**, 675–680 (1997).
7. Seielstad, M. T., Minch, E. & Cavalli-Sforza, L. L. *Nature Genet.* **20**, 278–280 (1998).
8. Wolpoff, M. & Caspari, R. *Race and Human Evolution* (Westview Press, Boulder, Colorado, 1997).

Why Spanish science is at a standstill

Sir — You reported the case of Antonio Ferriz Mas, a Spanish astrophysicist who has taken the University of Salamanca to court after being rejected for an associate professorship¹. Your editorial² made comments on the appointment process in Spanish universities and its alleged role as a significant obstacle to the development of science in Spain. You identified two main problems that prevent Spain's being influential in modern science: cronyism, the practice of favouring one's friends (internal candidates) in university appointments, and the “intellectual sclerosis” of a system based on tenured positions.

Support for the theory of cronyism comes from the fact that two of the five members of an appointment board come from the department or university involved. This suggests that social networks and unspoken agreements cause the selection, not of the most meritorious candidate, but of the one with the appropriate connections. A more adequate alternative might be to have panels that include just one member from the institution offering the position, or no-one at all. However, is it fair to limit so severely a university's contribution to decision-making on strategic issues which affect their long-term functioning, notably the appointment of tenured staff?

Decisions by any appointment board imply value judgements which are accepted, not because of their objectivity, but because their subjectivity is shared to a great extent by the scientific community. Common design is achieved through clear-cut, publicly visible criteria. Stability — permanence over time — is also a desirable criterion, especially when appointments are

made on the basis of long-term activities, as is usually the case in science. But even if these requirements are fulfilled, disagreements are still likely to occur. (See, for instance, the recent controversy over the exclusion of Salvador Moncada from the Nobel Prize³.) It is also plausible, as in any human activity, that from time to time regulations implemented to maintain fairness are overtaken and that biased decisions are intentionally made and adopted. If that is so in Ferriz Mas's case the court should say so. But to cast the slur of cronyism on the entire appointment procedure means casting doubts and allegations of corruption on the hundreds of university teachers who have participated in the process, either as panel members or as candidates. To us, that seems audacious, to say the least.

We could certainly debate the adequacy of the criteria followed by different appointment boards, and introduce improvements to the Spanish appointments system, not only at universities but at other research institutions. Certainly, it would be useful to establish some form of broad scientific/academic profile that young researchers can use as a reference for career planning. For university appointments, such a profile should, of necessity, include both research and teaching profiles, especially as tenure positions in Spain imply 240 hours' teaching per year, a significant amount of total labour time.

We fear, however, that even if this goal is achieved Spanish scientific development will still be at a standstill. Why? Because despite Spanish economic advances over the past years, investment in research and development, as a percentage of gross domestic product, is lower now than it was in 1991. It is the second lowest in the European Union, far below the EU average⁴, and even lower than in some eastern European states (such as the Czech Republic, Slovak Republic and Slovenia⁵). Furthermore, 43.8 per cent of university teachers, many of whom are highly qualified and experienced, face a dim future as most are under short-term non-tenured contracts⁶.

Meanwhile, postdocs like Ferriz Mas are sent abroad: 4,554 in the period 1984–94 (ref. 7). But, if the scientific structures and the political will at home remain the same, what's the point?

Angel Baltanás, Isabel Castro

Dept. de Ecología, Facultad de Ciencias, Universidad Autónoma de Madrid, E-28049 Madrid, Spain

1. Bosch, X. *Nature* **396**, 712 (1998).
2. *Nature* **396**, 709 (1998).
3. *Nature* **396**, 614 (1998).
4. Bosch, X. *Nature* **395**, 535 (1998).
5. Koenig, R. *Science* **283**, 22–24 (1998).
6. www.mec.es/estadistica/Cifras99/NAC_10.html
7. Martín-Sempere, M. J. et al. *Política Científica* **47**, 50–53 (1997).

Slow light in cool atoms

Jon Marangos

An experiment with atoms at nanokelvin temperatures has produced the remarkable observation of light pulses travelling at velocities of only 17 m s^{-1} . The large optical nonlinearities seen in this system may open up new opportunities in quantum optics.

In our usual understanding, the speed of light, c , is the absolute top speed in the Universe at $3 \times 10^8 \text{ m s}^{-1}$ in a vacuum. So observation of light pulses propagating at a speed no faster than a swiftly moving bicycle, described by Hau *et al.*¹ on page 594 of this issue, comes as a surprise. We know that light can be slowed to a modest extent in refractive and transparent media, for example water and glass, to velocities typically a factor of 1.5–2.0 times slower than c . But there is a limit to how much light can be slowed in normal optical materials, because the larger refractive index associated with slower propagation is inevitably accompanied by increased light absorption.

Under special circumstances, however, this limit can be overcome — that is, a perfectly transparent medium can be created in which the speed of light is slowed enormously. The systems in question are laser-dressed atomic media that acquire new optical properties because light does not interact directly with atoms but with a system composed of atoms plus laser field. This requires the preparation of laser-dressed atoms (see box for the technical details) to create what is termed electromagnetically induced transparency, in which quantum interference leads to the cancellation of absorption². In this new kind of system, the dispersive (or refractive) properties of the medium — including the velocity of propagation of an optical pulse — become independent of absorption³. For example, Steve Harris and colleagues⁴ at Stanford have used coherently prepared lead atoms to reduce the propagation velocity of a resonantly tuned light pulse to $c/165$. To achieve even slower pulse velocities, cold atoms are required because, to maximize the quantum interference effect, the thermal motion must be small.

Hau and her co-workers¹ use ultra-cold sodium atoms. They load a magneto-optical trap with sodium atoms, and the gas is cooled briefly with a laser to reach temperatures of $50 \text{ } \mu\text{K}$ (ref. 5). With the lasers switched off, only atoms in the ground state with magnetic dipoles directed opposite to the magnetic field are confined by the novel magnetic trap developed by the authors⁶. Hau *et al.* then evaporatively cool the atoms in this trap to reach temperatures in the

region of the Bose–Einstein condensation threshold $T_c = 435 \text{ nK}$. (Bose–Einstein condensates were first observed in 1995, in a famous experiment by Eric Cornell and Carl Wieman⁷, and are a unique state of matter in which all of the atoms exist in the same quantum state.)

In ultra-cold atoms, extremely narrow transparency dips due to quantum interference can be induced using very low powers of the ‘coupling’ laser beam. Accompanying this low absorption will be a very steep variation, with probe laser frequency, in the refractive index. This steep slope and the high sample density in the trapped cloud of atoms leads to ultra-slow light propagation. The velocity of the probe laser pulse increases with the coupling laser power and decreases

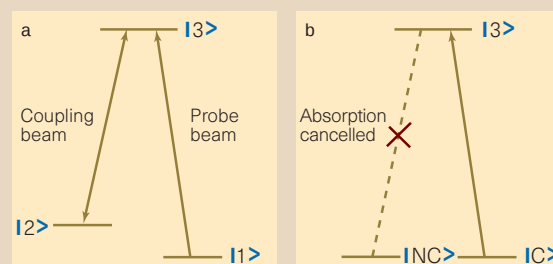
with atom density. So the lowest velocities are obtained by Hau *et al.* at low temperatures — for which strong interference can be seen with very low coupling laser power — and high sample density. The atom density shows an abrupt increase below T_c to values of $5 \times 10^{12} \text{ cm}^{-3}$, and although results for temperatures above T_c were obtained from the ultra-cold samples, the lowest velocities were always seen at temperatures below this value.

The coupling laser field and probe pulse used in this experiment (Fig. 1, overleaf) are derived from the same continuous-wave dye laser, with their frequency set by an acousto-optic modulator. The coupling laser is linearly polarized and directed perpendicular to the axis of the 0.2-mm-long cloud of atoms. The circularly polarized probe field is shaped into a $2.5\text{-}\mu\text{s}$ -long pulse by a second acousto-optic modulator. This pulse is propagated along the axis of the atom cloud. It takes the probe pulse about $7 \text{ } \mu\text{s}$ to pass through only 0.2 mm of cold atoms. Modest losses due to absorption are observed, although without quantum interference this medium would be completely opaque. The lowest velocity recorded was just 17 m s^{-1} at temperatures below T_c . For temperatures above T_c the velocities were up to a factor of four larger than this.

The reduction of the propagation velocity

Box 1: Laser-dressed atoms

In the experiment of Hau *et al.*¹, the sodium atoms can be thought of as a three-level atomic system subject to a pair of resonant laser fields. In **a**, the coupling field (linearly polarized) is applied to the unpopulated hyperfine states $|2\rangle$ and $|3\rangle$ and the probe pulse (circularly polarized) is applied to the $|1\rangle$ – $|3\rangle$ transition. In the absence of these fields the eigenstates of the system are simply those of the atomic Hamiltonian (that is, $|1\rangle$, $|2\rangle$ and $|3\rangle$). In **b**, in the presence of the fields the system has a new Hamiltonian of the atom plus laser field. Two of the eigenstates of this Hamiltonian ($|C\rangle$ and $|NC\rangle$) are a coherent superposition of $|1\rangle$ and $|2\rangle$. State $|C\rangle$ remains coupled to the fields, but can be ignored in this experiment as it is



unpopulated. In the populated eigenstate $|NC\rangle$ the probability amplitudes of each of the component states ($|1\rangle$ and $|2\rangle$) make equal but opposite contributions. This leads to a transition dipole moment from $|NC\rangle$ to $|3\rangle$ which exactly vanishes through destructive interference; this is electromagnetically induced transparency. This non-coupled state $|NC\rangle$ is called a ‘dark state’ to express how quantum interference has cancelled the interaction with the

laser fields¹⁰.

The quantum interference exists over a range of probe frequencies set by the coupling field power. In an ultra-cold sample complete absorption cancellation can be induced for very low coupling powers due to the very small magnitude of the atomic thermal motion. In this case there is a very narrow transparency dip and an abnormally steep dispersion profile that leads to a very slow pulse propagation velocity.

J. M.

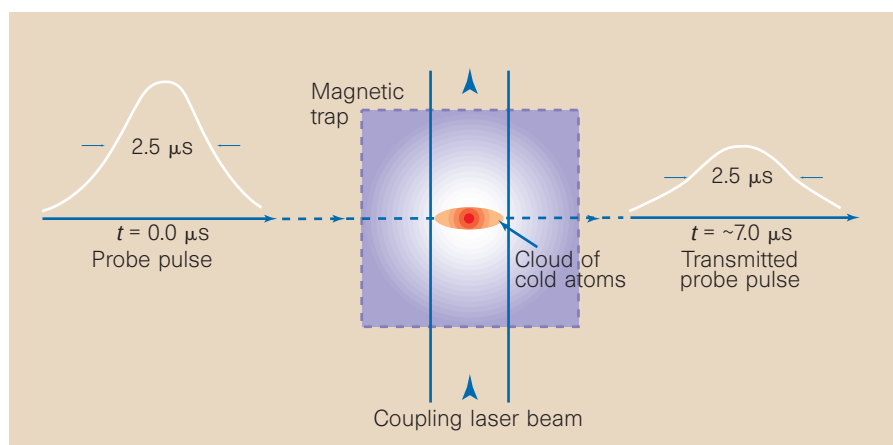


Figure 1 The experimental set-up of Hau *et al.*¹. The circularly polarized probe pulse takes about $7 \mu\text{s}$ to traverse the 0.2-mm cloud of cold atoms — this is $1/10$ millionth of the speed of light in a vacuum.

ty of the probe pulse — which is by a factor of about 20 million compared with the speed of light in a vacuum — is a stunning proof of the dramatic changes that can occur to the optical properties of laser-dressed atoms. In contrast to a normal atomic medium, the atoms here are in the presence of the coupling laser field so the probe pulse should be thought of as interacting with a system composed of the atoms plus the coupling field. The atoms alone could not have stored the energy of the probe pulse for the required time because of the normal dissipative process of spontaneous emission (see box). However, when the probe pulse enters the medium, its energy goes into the combined atom and field system where it is immune from rapid spontaneous decay. At the end of transmission the energy is returned to the probe field from the system. The energy of the probe pulse is, in a sense, kept safe within the non-decaying 'dark state' of the atom created by quantum interference.

Hau *et al.* have not yet found out if the atom cloud remains in the Bose–Einstein condensed state during the interaction with the probe. It may be that the main role the condensate plays in the slow velocity propagation is in providing a high density of cold atoms. The authors propose that with some technical improvements (higher frequency stability, lower coupling powers) still lower velocities can be achieved, perhaps down to a few centimetres per second.

Another important aspect of the work was the observation of large optical nonlinearities, in the form of an intensity-dependent refractive index. This was inferred from measuring the intensity-dependent frequency shift in the position of the transparency peak. Hau *et al.* conclude that, at $0.18 \text{ cm}^2 \text{ W}^{-1}$, the nonlinear refractive index is unprecedentedly large.

Earlier work also considered potential applications of the optical properties of laser-dressed atoms. For instance, there have been detailed proposals to use these systems in highly sensitive magnetometers⁸ or as an

intracavity element to narrow a laser cavity linewidth. The slow pulse velocities reported by Hau *et al.* have yet to find a specific application, but laser-dressing clearly results in profound modifications of the optical properties of the medium. In the laser-dressed cold atoms, the pulse takes $7 \mu\text{s}$ to traverse the 0.2-mm sample. Propagating in a vacuum for the same period, by contrast, the pulse would have travelled 2 km ! Perhaps this phenomenon could be used in optical delay lines for generating very long delays, or in allowing a shorter reference arm for an interferometer in which the other arm is many kilometres in length. Other applications might use the long

time the pulse can be stored in the medium without significant dissipation, for instance in optical data storage.

Finally, the massive nonlinearities observed in this system are of a type that lead to a strong coupling between pairs of photons. Photons are particles that normally cannot interact strongly, so this is an unusual regime. Potentially these interactions may be large enough that it would become possible for a single photon to switch an optical cavity⁹. Nonlinearities of this kind have also been shown to be the key ingredient in experiments in quantum optics such as optical squeezing, quantum non-demolition measurements and studies of non-locality. The increased magnitude in the nonlinearities observed by Hau *et al.* may lead to improvements in these experiments. □

Jon Marangos is in the Laser Optics and Spectroscopy Group, The Blackett Laboratory, Imperial College of Science, Technology and Medicine, London SW7 2BZ, UK.

e-mail: j.marangos@ic.ac.uk

1. Hau, L. V., Harris, S. E., Dutton, Z. & Behroozi, C. H. *Nature* **397**, 594–598 (1999).
2. Harris, S. E. *Phys. Today* **50**, 36–40 (1997).
3. Scully, M. O. *Phys. Reps* **219**, 191–201 (1992).
4. Kasapi, A. *et al.* *Phys. Rev. Lett.* **74**, 2447–2450 (1995).
5. Chu, S. *et al.* *Rev. Mod. Phys.* **70**, 685–741 (1998).
6. Hau, L. V. *et al.* *Phys. Rev. A* **58**, R54–R57 (1998).
7. Anderson, M. H. *et al.* *Science* **269**, 198–201 (1995).
8. Scully, M. O. *et al.* *Phys. Rev. Lett.* **69**, 1360–1363 (1992).
9. Yamamoto, Y. *Nature* **390**, 17–18 (1997).
10. Arimondo, E. *Prog. Opt.* **35**, 257–354 (1996).

Planetary science

Snapshots of an ancient cover-up

Maria T. Zuber

Mars is a superb and arguably unique natural laboratory for the study of climate in the early stage of evolution of the terrestrial planets. As described in four papers in this issue^{1–4}, beginning on page 584, spectacular images from the Mars Orbiter Camera, an instrument on the Mars Global Surveyor spacecraft currently orbiting Mars, are providing a new view of the processes that shaped the Martian surface.

Shortly after the planets formed 4.5 billion years ago, planetesimal impacts on their surfaces, and the differentiation of metal and rock to form planetary cores and mantles, heated the interiors of the terrestrial planets. As these planets rapidly cooled they liberated volatile gases to form atmospheres and, at least on Earth, the oceans. Both Earth and Mars underwent these processes, but the geological record on Earth predominantly preserves surfaces younger than 500 million years old; earlier records have been eradicated by processes such as subduction, continental collision and erosion. On Mars, however, most surfaces have ages greater than three billion years and they retain tantalizing

hints that the early heat-loss phase on this planet produced climatic conditions that were much more hospitable than the cold, desert-like environment that exists now^{5,6}.

Today, Mars is too cold and the atmosphere is too thin for liquid water to be stable at the surface. Most water is currently stored as ice in the polar caps and in 'frozen aquifers' beneath the surface. However near-global images, of moderate resolution (around 200 metres per pixel) that were taken by the Mariner 9 and Viking Orbiter space probes in the 1970s, revealed a planet on which liquid water had flowed on the surface. From the early images, a diversity of early Martian climates were proposed, ranging from warm conditions with hemispheric-scale oceans to near-freezing conditions where water flowed on the surface for only the briefest of periods. Previous data have not been able to distinguish between specific evolutionary schemes, but spectacular new high-resolution images from the Mars Orbiter Camera (MOC) are allowing the range of possibilities to be narrowed considerably. In the four reports^{1–4} now published, MOC principal

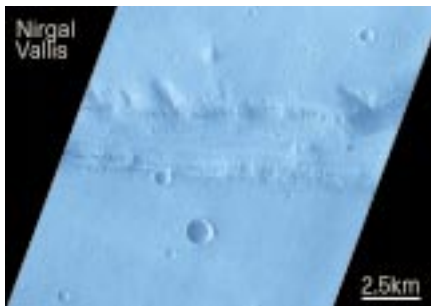


Figure 1 Mars Orbiter Camera image 605 illustrates, in a single view, many of the insights about Mars gained from recent high-resolution pictures¹⁻⁴. It shows Nirgal Vallis, one of the Martian valley systems, and the surrounding surface. The absence of tributaries from the surface of the plains, and the presence of side-wall alcoves, point to its origin through groundwater processes rather than from precipitation-fed run-off. The floor of the canyon is filled with sediments that form dunes, obscuring features that might have formed by sustained flow within the canyon. On the upland, surface craters in a range of states of preservation attest to a dynamic history of cratering and atmospheric dust deposition that are likely to obscure age relationships.

investigator Michael Malin and colleagues analyse images with a resolution of four to eight metres per pixel — 20 to 40 times better than obtained before.

Malin and Carr³ (page 589) address the origin of valley networks, which are kilometre-wide channels that occur on some of the oldest (typically 3.5-billion-year-old) surfaces on the planet. Whereas previous orbiters imaged these features, which represent evidence for flowing water⁵, it has been debated whether their source was predominantly precipitation or ground water^{7,8}. Malin and Carr show images of channels that clearly indicate a groundwater source (Fig. 1). In addition, an image of a meandering valley in Nanedi Vallis indicates that water, at least in this area, flowed on the surface for an extended period rather than being due to a transient catastrophic outflow. Other images show that wind-distributed material has filled certain channels, in some cases completely, probably obscuring additional evidence for early surface water.

The report by Thomas *et al.*⁴ (page 592) specifically describes the modification of the Martian surface by windblown dust. These authors describe surprising evidence that Martian dunes are composed of two distinctive materials, demonstrating that dust on Mars has not completely homogenized over geological history, despite mixing by global dust storms. Of particular significance is the identification of widespread deposits of sediment up to thicknesses of tens of metres. As Hartmann *et al.*² discuss (page 586), sediments are often deep enough to fill small craters completely, which is problematic

because records of craters of various sizes are used to determine the relative ages of planetary surfaces. The new data exacerbate a previously recognized problem⁹ — that the ages of some Martian surfaces may have been underestimated — making it difficult for scientists wishing to construct a consistent timeline of geological and environmental events. In addition, the accumulation of thick sediments presents a challenge to future NASA lander missions that want to sample ancient bedrock.

Finally, on page 584, McEwen *et al.*¹ provide evidence for ubiquitous, fine-scale layering within the vast Valles Marineris canyon. Such detailed wall structures had not been previously observed in any of the Martian canyons. From the few images obtained so far it is not possible to unambiguously distinguish between a volcanic or sedimentary origin for the layering, but this may be resolved with upcoming observations. In either case, the existence of the layering is significant.

Previous studies have suggested that the dominant process for shaping the earliest surfaces was pervasive fracturing and mixing of crustal material by impacts. But it is now apparent that volcanic and/or sedimentary processes operated extensively during this early period. If the origin of the layering is sedimentary it should be possible to address the questions of the volume and persistence of surface water needed to deposit the observed 8-km-thick column. A volcanic origin would imply the release to the surface of significant amounts of subsurface water, and would allow for better estimates of the amount of water released in outflow channels and valleys. In addition, volcanic outgassing represents a source of atmospheric gases, and is a dominant factor in determining whether Mars ever developed a runaway greenhouse effect¹⁰, which has been invoked in climate

theories favouring a 'warm, wet' early Mars.

The picture that emerges from these high-resolution views is unmistakable evidence of the dynamical processes that shaped primordial Mars. Unfortunately, the processes that appear to have dominated early Martian history have also conspired to cover up surfaces of all ages and may be obscuring evidence of Mars' early climate. But the images obtained so far by Malin and colleagues cover only a minuscule fraction of the Martian surface and provide only enticing teasers of observations to come. Orders of magnitude increases in spatial sampling to several metres resolution are expected from the MOC during the two-year-long continuous mapping phase of Mars Global Surveyor, which commences in March. These images, combined with observations from other orbital sensors and with local views from surface landers, offer the hope of piecing together a self-consistent view of Mars' early climate — with implications for the development, or not, of early life on the planet. □

Maria T. Zuber is in the Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA.

e-mail: zuber@mit.edu

1. McEwen, A. S., Malin, M. C., Carr, M. H. & Hartmann, W. K. *Nature* **397**, 584–586 (1999).
2. Hartmann, W. K. *et al.* *Nature* **397**, 586–589 (1999).
3. Malin, M. C. & Carr, M. H. *Nature* **397**, 589–591 (1999).
4. Thomas, P. C. *et al.* *Nature* **397**, 592–594 (1999).
5. Carr, M. H. *Water on Mars* (Oxford Univ. Press, New York, 1996).
6. Squyres, S. W. & Kasting, J. F. *Science* **265**, 744–748 (1994).
7. Gulick, V. C. & Baker, V. R. *Nature* **341**, 514–516 (1989).
8. Baker, V. R. *et al.* in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 493–552 (Univ. Arizona Press, Tucson, 1992).
9. Chapman, C. R. & Jones, K. L. *Rev. Earth Planet. Sci. Lett.* **5**, 515–540 (1977).
10. Fanale, F. P. *et al.* in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 1135–1179 (Univ. Arizona Press, Tucson, 1992).

Neurobiology

A place for space and smells

J. N. P. Rawlins

The hippocampus lies deep within the temporal lobes of the human brain, and clinical studies suggest that it is essential for normal, everyday memories of many different events. In contrast, animal studies indicate that the hippocampus is a highly specialized memory store, dealing with a strictly limited information set. This issue has been hotly debated over the past two decades or so and, on page 613 of this issue, Wood *et al.*¹ report data from a new animal study, which adds weight to the view that the hippocampus has a general role in memory.

The hippocampus has a simple, cortical structure in which the neurons are arranged

in three layers (rather than the six layers of the neocortex, which makes up the convoluted surface of the brain). The name hippocampus derives from its elegantly curved structure, somewhat fancifully supposed to resemble the shape of a seahorse. (An alternative name, Ammon's horn, reflects a more plausible similarity to the horns of a ram-headed Egyptian god.)

Experimental analysis of hippocampal function was kick-started by clinical observations in the early 1950s. A patient known as H. M., who suffered from uncontrollable epilepsy, was treated with a "frankly experimental" surgery — the structures on the inner faces of his left and right temporal

lobes were removed. The surgery led to a devastating amnesia and, although H. M. could still recall events from before the operation, he seemed unable to lay down new memories, retaining memories of new events for no more than a few minutes². In the hunt to identify which component of H. M.'s brain damage might be responsible for this striking clinical outcome, attention focused on the hippocampus, which was seriously damaged by the surgery.

Interest increased still further when, in the mid-1970s, long-term potentiation was described in the hippocampus³. This was a new electrophysiological demonstration that the ability of neurons to activate one another could be adjusted as a function of experience. The discovery placed a plausible mechanism for memory storage within a structure whose damage was associated with profound, clinical memory loss. It seemed increasingly likely that the hippocampus was critical for memory — but what kind of memory?

The most influential theory of hippocampal function remains the spatial-mapping theory, put forward by O'Keefe and Nadel⁴ in 1978. They suggested that the hippocampus acts exclusively to create and store maps of space. A cognitive map of this kind would store information about the relative spatial positions of important goals (such as the location of food, water or nests) and other objects. From this, animals could then derive the best path from one location to another (and calculate alternative paths if the familiar route was blocked).

Some of the evidence supporting this theory comes from experimental lesion studies. Spatial tasks can be impaired by hippocampal damage, whereas otherwise equivalent, non-spatial tasks are spared. For example, rats with hippocampal damage cannot learn how to swim directly to a specific place in a pool where an escape platform is hidden just under the water. However, the rats rapidly learn to find the platform if there is some distinctive, visible cue just above it⁵. Normal rats, on the other hand, easily learn both versions of the task. But perhaps the most compelling support for the spatial-mapping theory comes from single-neuron recording studies done in freely moving animals. These studies have concluded that spatial location is a major determinant of neuronal activity in the hippocampus.

Evidence from the early 1970s onwards has shown the existence of 'place fields' in hippocampal neurons⁶. This means that the neurons fire whenever an animal is in a particular place. 'Places' seem to be defined by their relation to combinations of cues in the surrounding environment (doors, windows, pictures, sources of noise and smells, and so on), but do not depend on any one of these cues alone. So, individual cues can often be removed without losing the neuron's place

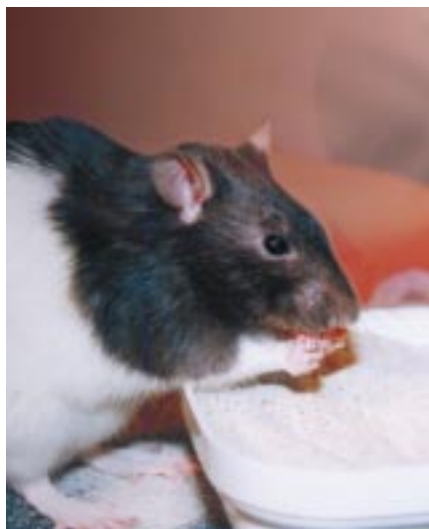


Figure 1 Smells good, tastes good — successful completion of the odour-memory task set by Wood *et al.*¹.

field, as long as enough of the other cues remain. Such observations indicate that relational, spatial cues are critical for hippocampal activity, consistent with a specialized function for the hippocampus in cognitive mapping.

Wood *et al.*¹ now challenge that conclusion. They recorded the activity of hippocampal neurons in rats performing an odour-memory task on an apparatus surrounded by a rich set of environmental cues. The rats learned to dig for food in small cups of scented sand. Food was found only in cups that smelt different from the last one the rat had dug in. The only way to avoid digging fruitlessly was by comparing the scent of the present cup with the remembered scent of the previous cup. So, the rats had to depend on a non-spatial olfactory memory to do the task efficiently.

Wood and colleagues varied the spatial locations of the cups, the specific odours, and whether or not the scent of one cup matched that of the preceding cup. Under these conditions, the authors not only found neurons that fired in particular spatial locations, but they also found neurons whose activities depended on non-spatial variables — some responded to particular odours, others fired when the present cup smelled like the previous cup, but did not fire when the scents didn't match. Other cells responded to particular combinations of properties.

Previous studies have reported that hippocampal cells can respond to non-spatial stimuli⁷. The key finding by Wood *et al.* is not that the cells did this, but that they were slightly more likely to have non-spatial than spatial determinants (even though the authors used the same kind of rich spatial environment typically used to study the formation of place fields). Furthermore, activity of the cells in the presence or absence of

their particular determinants was similar, irrespective of whether those determinants were spatial. These results indicate that hippocampal neurons are equally available to process spatial and non-spatial data, and to much the same extent. Previous studies showing a clear preponderance of spatially sensitive neurons could reflect experimental design rather than a specialization for spatial processing in the hippocampus. The outcomes have, perhaps, told us more about the minds of experimenters than about the brains of rats.

Where next? Recording studies⁸ have shown that neurons with a place field in one environment can also have place fields in other environments. Can specific neurons with well-defined, non-spatial determinants in one situation be redeployed to support a spatial function with a well-defined place field in another situation? This would be striking support for a general-memory theory of hippocampal function. And what happens to behaviour if these units with non-spatial determinants are unavailable? In some studies, hippocampal damage does not impair memory for small objects under conditions where it impairs memory for larger, spatial contexts⁹. Would hippocampal dysfunction disrupt the performance of rats in Wood and colleagues' non-spatial odour-memory task in the same way that it disrupts performance of spatial tasks? If not, the conclusion would be that hippocampal neurons may have no direct, active role in the non-spatial processing with which they are correlated.

With their study, Wood *et al.*¹ add weight to the proposal that the functions of the hippocampus cannot be subsumed within a purely spatial memory framework^{10,11}. This may push theories of hippocampal memory derived from single-unit studies back towards theories derived from studies of amnesic humans. And it should certainly drive further experimental comparisons of hippocampal functioning in spatial and non-spatial tasks. □

J. N. P. Rawlins is in the Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK.
e-mail: nicholas.rawlins@psy.ox.ac.uk

1. Wood, E. R., Dudchenko, P. A. & Eichenbaum, H. *Nature* **397**, 613–616 (1999).
2. Penfield, W. & Milner, B. *Arch. Neurol. Psychiatry (Lond.)* **79**, 475–497 (1958).
3. Bliss, T. V. P. & Lomo, T. J. *Physiol. (Lond.)* **232**, 331–356 (1973).
4. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Oxford Univ. Press, 1978).
5. Morris, R. G. M., Garrud, P., Rawlins, J. N. P. & O'Keefe, J. *Nature* **297**, 681–683 (1982).
6. O'Keefe, J. & Dostrovsky, J. *Brain Res.* **34**, 171–175 (1971).
7. Berger, T. W., Rinaldi, P. C., Weisz, D. J. & Thompson, R. F. *J. Neurophysiol.* **50**, 1197–1219 (1983).
8. Müller, R. U. & Kubie, J. L. *J. Neurosci.* **7**, 1951–1968 (1987).
9. Cassaday, H. J. & Rawlins, J. N. P. *Behav. Neurosci.* **111**, 1228–1244 (1997).
10. Rawlins, J. N. P. *Behav. Brain Sci.* **8**, 479–528 (1985).
11. Eichenbaum, H., Otto, T. & Cohen, N. J. *Behav. Brain Sci.* **17**, 449–518 (1994).

Palaeontology

Something fishy in the family tree

Per Erik Ahlberg

Where did the bony fishes come from? This simple question illustrates a theme, the quest for origins, which has been central to evolutionary biology since its birth as a discipline. Great effort has been expended on tracing the beginnings of the major groups of organisms in the fossil record. This research has had its successes, but has been curiously frustrating — time and again, the vaunted 'basal' forms it has uncovered have confounded expectations, leaving us as much puzzled as enlightened. Now, Zhu *et al.* (page 607 of this issue¹) uncover another example of this pattern: they present a basal bony fish with such an unexpected mix of characteristics that it forces a reconsideration of large parts of the vertebrate family tree.

The 'bony fishes', Osteichthyes, a group which in fact also includes the land vertebrates or tetrapods, and thus ourselves, is by far the largest group of backboned animals. It musters some 50,000 species, as against roughly 700 sharks, rays and chimaeras, and a mere handful of jawless lampreys and hagfishes. The group's ecological importance is incalculable. It might seem odd to unite animals as different as a cod (a decent bony fish by any measure) and a human being in the same group, but all osteichthyans have a suite of common features which show beyond doubt that they share a single common ancestry. Tetrapods have evolved from more orthodox bony fishes². Clearly, the origin of the osteichthyans was a momentous event in the history of life, but attempts to investigate this event always come up against a baffling barrier.

There is almost universal agreement that the Osteichthyes can be divided into two branches, the Actinopterygii or ray-fins and the Sarcopterygii or lobe-fins (Fig. 1). The actinopterygians, which include the cod and perhaps 25,000 other species, form the largest living fish group. The sarcopterygians include the phenomenally successful land vertebrates, but only four living genera of fishes — the Australian, African and South American lungfishes, and the coelacanth *Latimeria*.

However, as you trace these groups backwards in time through the fossil record, sarcopterygian fishes become more diverse until, in the Devonian Period, 408 to 363 million years ago, we find a wide range of lobe-fins alongside just a few ray-fins. But then the record stops, abruptly; no Silurian rocks have ever yielded more than isolated scales and scraps of osteichthyans, and in older deposits they are entirely absent. The Silurian forms have been given names, *Lophosteus*

and *Andreolepis*, but we know almost nothing about them and cannot tell how they relate to later osteichthyans.

Phylogenetic investigations, whether based on fossils, molecular data or the morphology of recent osteichthyans, tend to hit the buffers in a similar way. They generally support the actinopterygian–sarcopterygian split, but have failed to resolve unequivocally how the main sarcopterygian groups are related to each other^{3–6}. Furthermore, while almost invariably placing the Chondrichthyes (sharks, rays and chimaeras) as the living sister group of the Osteichthyes, they have not established which of the fossil non-osteichthyan groups is closest to the Osteichthyes⁷. As a result of these limitations of phylogenetic resolution and the fossil record, we don't know what the last common ancestor of the Sarcopterygii, or of the Osteichthyes as a whole, was like; we don't know which characteristics are primitive for the different subgroups; and we cannot tell from what kind of ancestry the Osteichthyes emerged.

Into this confusion drops Zhu and colleagues' discovery, the 400-million-year-old *Psarolepis* from the earliest Devonian and latest Silurian of Yunnan in south China^{1,4,8}. It is the earliest osteichthyan known from

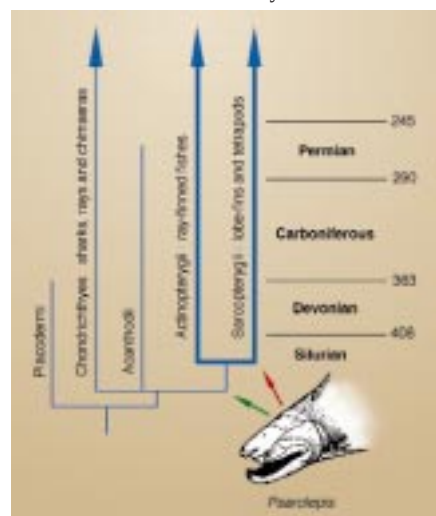


Figure 1 Fish relationships and the ambiguous position of *Psarolepis*. This phylogeny shows the two positions of *Psarolepis* obtained by Zhu *et al.*¹ using the data sets of Cloutier and Ahlberg³, and Zhu and Schultze⁴. The first data set places *Psarolepis* as a stem sarcopterygian (red arrow), the second as a stem osteichthyan (green arrow). The osteichthyan groups are shown in bold; arrowheads at the tops of lines indicate groups persisting to today, numbers indicate ages in millions of years. Drawing of *Psarolepis* from ref. 1.

reasonably complete remains of skull and shoulder girdle, and shows a strikingly incongruous suite of features. The braincase, which is divided into separate front and back halves by a joint, looks like that of a reasonably orthodox sarcopterygian; so do the teeth, which have a characteristically sarcopterygian type of folded 'polyplocodont' dentine. Three conspicuous openings on the outer face of the lower jaw, and a matching set on the cheek, again point to sarcopterygian affinities, specifically with a group of supposed lungfish relatives known as porolepiforms.

But the tooth-bearing bones of the snout and lower jaw match those of actinopterygians, as do the cheek bones. The shoulder girdle doesn't even look osteichthyan; a large spine sweeps outwards just in front of the pectoral fin, much as in an extinct group of fishes called placoderms. This is interesting, because it has been argued that placoderms are the closest relatives of osteichthyans⁷. However, *Psarolepis* also seems to have spines in front of its median fins, a feature more reminiscent of another group of putative osteichthyan relatives, the acanthodians⁷.

So where does all this leave us? Clearly, existing ideas about what features characterize the different osteichthyan groups are much too simplistic. Some of the supposedly unique 'actinopterygian' and 'sarcopterygian' characteristics are probably primitive for the whole Osteichthyes, and their presence side-by-side in *Psarolepis* is simply indicative of the primitive nature of this fish, while the placoderm- and acanthodian-like features may be previously unrecognized components of the primitive osteichthyan character complex. (At least one such feature, the pectoral spine, is also present in the Silurian genus *Andreolepis*¹.)

These re-evaluations will in due course have a major influence on the interpretation of osteichthyan phylogeny. For the moment, Zhu *et al.*¹ try to establish a preliminary phylogenetic position for *Psarolepis* by incorporating it, together with an acanthodian, a placoderm and a shark, into two competing published osteichthyan data sets^{3,4}. The results are interesting, but also frustrating (Fig. 1); depending on which data set is used, *Psarolepis* becomes either a stem osteichthyan or a stem sarcopterygian, and the judgements about character polarities and distributions differ accordingly. It seems we are still trapped by existing phylogenetic uncertainties.

In the long term, however, the prospects look much brighter. First, the existing data sets need to be revised (I am the co-author of one of them³, and would certainly not regard it as a final statement). Second, a number of other early and/or primitive osteichthyans, including several more from Yunnan, are currently under study and will be described

within the next few years. These should help to establish whether *Psarolepis* really is incongruous, or whether its combination of features is typical of the earliest osteichthyans. □

Per Erik Ahlberg is in the Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK.
e-mail: pea@nhm.ac.uk

1. Zhu, M., Yu, X. & Janvier, P. *Nature* **397**, 607–610 (1999).
2. Ahlberg, P. E. & Milner, A. R. *Nature* **368**, 507–514 (1994).
3. Cloutier, R. & Ahlberg, P. E. in *Interrelationships of Fishes* (eds Stiassny, M. L. J., Parenti, L. R. & Johnson, G. D.) 445–479 (Academic, San Diego, 1996).
4. Zhu, M. & Schultze, H.-P. *Lethaia* **30**, 293–304 (1997).
5. Gorr, T., Kleinschmidt, T. & Fricke, H. *Nature* **351**, 394–395 (1991).
6. Zardoya, R. & Meyer, A. *Proc. Natl Acad. Sci. USA* **93**, 5449–5454 (1996).
7. Janvier, P. *Early Vertebrates* (Oxford Univ. Press, 1996).
8. Yu, X. *J. Vert. Paleontol.* **18**, 261–274 (1998).

Physical chemistry

The protean proton in water

James T. Hynes

The character and anomalously high mobility of the hydrated proton in liquid water have long exercised researchers, providing inspiration for widely disparate and conflicting theories¹. Recent years have seen an intensification of this effort, especially on the theoretical front, and on page 601 of this issue² Marx *et al.* address these questions using perhaps the most powerful computer simulation techniques to date. The authors conclude, as have others, that the most widely held views of the structure and transport of the hydrated proton should be replaced by a more complex and rich picture, in which the proton's quantum-mechanical character plays an important role. These results should have an impact in areas related to aqueous acid–base chemistry³, enzymatic proton transfers⁴ (including 'low-barrier hydrogen bonds'⁵), proton transport in biological channels ('proton wires')⁶ and atmospheric proton transfers on ice surfaces⁷.

To most chemists, the hydrated proton is H_3O^+ , the oxonium ion (the attraction of the bare proton to a water molecule is so large that the proton could not exist in isolation). This picture was refined in the 1960s by Eigen⁸, who advocated the larger complex H_9O_4^+ , which is the current prevailing view. In this 'Eigen cation', the central oxonium ion is strongly hydrogen-bonded to three H_2O molecules — that is, it is in the form $(\text{H}_3\text{O}^+)(\text{H}_2\text{O})_3$ — with interactions with remaining water molecules being weaker.

But it is not immediately obvious that the Eigen cation should be the dominant hydrated-proton structure in liquid water. The protonated gas-phase water dimer H_5O_2^+ is evidently $\text{H}_2\text{O} \cdots \text{H}^+ \cdots \text{H}_2\text{O}$, with a minimum energy configuration which places the proton midway between the two water molecules⁹. But because of the proton's small mass, its quantum zero-point energy is significant, so that the proton does not actually reside at this minimum, but instead is delocalized to a degree in-between the flanking water molecules. To argue that the relevant solution structure is the Eigen cation would involve the surrounding polar water molecules preferring a more localized charge distribution

— here the H_3O^+ core — rather than the more delocalized H_5O_2^+ charge distribution. Such aspects already highlight essential features of the hydrated proton: its electronic structure depends strongly on the surrounding water molecules, and quantum nuclear issues are involved.

There is, however, a contrasting and until recently minority view attributed to Zundel¹⁰, namely that the hydrated proton is in fact H_5O_2^+ , the 'Zundel cation'. An important piece of evidence is the striking infrared 'continuum' observed in concentrated acid solutions, indicative of a wide and continuous range of the vibrational energy of the hydrated proton. In the H_5O_2^+ scheme, this arises from the large proton polarizability within the cation; the fluctuating electric fields from the surrounding water molecules would then have a large variable effect on the proton's vibrational levels, producing the continuum¹⁰.

Intimately related to the structural question is the mechanism of the anomalously high mobility of the proton in water — approximately five times that of ions of similar size to H_3O^+ . This has been explained using the famous Grotthuss mechanism¹. Broadly speaking, this involves a proton shuttling between successive water molecules, and, rather than involving the larger H_3O^+ or Eigen cation transport, a proton relay in the hydrogen-bonded water network takes place. Here, however, general agreement ends, and many divergent mechanisms have been proposed¹.

In 1964 Luz and Meiboom¹¹ measured a very short timescale of about 1 picosecond for aqueous proton transport. This was an early signal that the hydrated proton 'structure' must be a delicate thing, with rapidly interconverting transient structures being involved¹². In addition to the difficulties it poses to experimentalists, this apparently low activation barrier increases the complexity of electronic structure calculations, which must be addressed in the milieu of the strongly interacting water solvent. Together with the need to account for the proton's quantum-mechanical character, these features make it clear that the hydrated proton

presents a significant challenge to scientists trying to unravel the process.

Marx *et al.*² address the above issues using a powerful computer simulation methodology that includes both the quantum-electronic and quantum-nuclear motion. This *ab initio* molecular dynamics approach differs fundamentally from the current workhorse of classical molecular dynamics. The latter uses Newton's classical laws of motion, and employs forces whose form is predetermined empirically or by separate electronic structure calculations. The former instead determines the forces 'on the fly' by way of electronic structure calculations during the course of the dynamics, while the nuclear quantum wave character is handled using Feynman path integral techniques.

Marx and colleagues' Fig. 2a (page 603) indicates a central result: the hydrated proton is of an inherently fluxional character, with the defined Eigen and Zundel cations having a comparable degree of incidence, but with many other intermediate structures occurring in the observations. The comparison with classical mechanical calculations (Fig. 2b) clearly emphasizes the importance of the delocalized proton's quantum zero-point energy, which serves to enhance its essentially fluxional, protean character.

The authors also find that aspects of the proton transport in their model disagree with many traditional views. One crucial feature is that, in the hydrated proton's interconversion through its assorted forms, the rate is not limited by the proton motion itself, but rather involves water solvent rearrangement. For instance, the breaking of a hydrogen bond between the proton-accepting water molecule and a nearby water molecule 'prepares' the proton-accepting water for its future identity as an H_3O^+ core. The light, fast proton then follows such rearrangements in the aqueous surroundings, which are slower.

In view of the delicate and complex — not to mention controversial — character of the hydrated proton, further detailed exploration can clearly be expected through this and other^{13–15} approaches, together with experimental measurement of the structure, rate and spectroscopy. Beyond this, exciting applications to the outstanding problem areas mentioned at the beginning of this article are surely on the horizon. □

James T. Hynes is in the Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309, USA.

e-mail: hynes@spot.colorado.edu

1. Agmon, N. *Chem. Phys. Lett.* **244**, 456–462 (1995).
2. Marx, D., Tuckerman, M. E., Hutter, J. & Parrinello, M. *Nature* **397**, 601–604 (1999).
3. Ando, J. & Hynes, J. T. *J. Mol. Liq.* **64**, 25–37 (1995).
4. Kohen, A. & Klinman, J. P. *Acc. Chem. Res.* **31**, 397–404 (1998).
5. Cleland, W. W. & Kreevoy, M. M. *Science* **264**, 1887–1890 (1994).

6. Pomes, R. & Roux, B. *Chem. Phys. Lett.* **234**, 416–424 (1995).
7. Bianco, R. & Hynes, J. J. *Phys. Chem. A* **102**, 309–314 (1998).
8. Eigen, M., Kruse, W. & De Maeyer, L. *Prog. React. Kin.* **2**, 285–318 (1964).
9. Kochanski, E., Kelterbaum, R., Klein, S., Rohmer, M. M. & Rahmouni, A. *Adv. Quantum Chem.* **28**, 273–291 (1997).
10. Zundel, G. in *The Hydrogen Bond – Recent Developments in Theory and Experiments II* (eds Schuster, P., Zundel, G. & Sandorfy, C.) 683–766 (North Holland, Amsterdam, 1976).
11. Luz, Z. & Meiboom, S. *J. Am. Chem. Soc.* **86**, 4768–4769 (1964).
12. Newton, M. J. *Chem. Phys.* **67**, 5535–5546 (1977).
13. Vuilleumier, R. & Borgis, D. J. *Phys. Chem. B* **102**, 4261–4264 (1998).
14. Schmitt, U. W. & Voth, G. A. *J. Phys. Chem. B* **102**, 5547–5551 (1998).
15. Limbach, H.-H. & Manz, J. (eds) *Ber. Bunsenges. Phys. Chem.* **102**, 289–586 (1998).

Synthetic chemistry

Sugaring vancomycin

Dudley H. Williams

In a recent issue of *Angewandte Chemie International Edition*, Nicolaou and co-workers¹ announced the total synthesis of the clinically important antibiotic vancomycin. The details of the final steps of this synthesis might turn out to be significant in the quest for modified antibiotics, which will, one hopes, perform better against serious clinical pathogens.

The discovery and exploitation of penicillin in the 1930s was a crucial advance in fighting bacterial infections, and was especially important during the Second World War. But, as early as the late 1950s, many strains of the bacterial pathogen *Staphylococcus aureus* were resistant to penicillin. The problem was initially kept at bay with the help of a semi-synthetic penicillin (methicillin). But methicillin-resistant *S. aureus* (MRSA) soon appeared and, by the 1980s, became resistant to the vast majority of available antibiotics². Resistance arises from the process of Darwinian selection — antibiotics wipe out susceptible strains, but those chance mutants that possess antibiotic resistance survive.

MRSA is the 'superbug' of the popular

press; it occasionally causes the closure of a hospital ward or an operating theatre, and is often a serious problem in hospitals where antibiotics are used intensively². The antibiotics of last resort against MRSA in hospitals are vancomycin and teicoplanin — both members of the glycopeptide family. These antibiotics (with combined sales in the region of US\$1 billion) were brought to the clinic through the efforts of scientists at Eli Lilly (Indianapolis) and Lepetit (Milan). They act by binding to growing bacterial cell-wall components³, and so are able to prevent the production of viable bacteria. Inevitably, resistance to vancomycin is now emerging in some pathogens, and a particular problem is associated with vancomycin-resistant enterococci (VRE)^{4,5}.

So, it is perhaps not surprising that chemists have aimed for a total synthesis of vancomycin. Not only is it a molecule of wide interest and enormous importance, but its structure poses a remarkable challenge to the synthetic chemist. Vancomycin is made up of seven amino acids, joined together to give a heptapeptide (Fig. 1). The side chains of several of the amino acids are

crosslinked, resulting in a closely packed structure. The rings, which are formed by crosslinking, are sufficiently small that the chlorine atom of the side chain of amino-acid 2 is fixed (at room temperature) at the 'back' of the molecule: the benzene ring to which it is attached is prevented from rotating (see arrow 'a' in Fig. 1) by the surrounding atoms. Similarly, the chlorine atom attached to the side chain of amino-acid 5 is constrained at the 'front' of the molecule. The synthetic chemist has to make sure that, in the final product, the chlorine atoms 'point' in the appropriate direction. A further complication is that the crosslinking of the side chains of amino-acids 5 and 7 must be achieved so that the OH group attached to the benzene ring of amino-acid 5 is more remote from the viewer of Fig. 1 than the benzene ring of amino-acid 7. That is, a 180° rotation about the bond linking the benzene rings of amino-acids 5 and 7 cannot occur in the synthetic vancomycin (see arrow 'b' in Fig. 1). Moreover, all the amino-acid constituents of vancomycin must first be synthesized with the appropriate stereochemistry, and fastened together to give the heptapeptide backbone (which runs through the centre of the molecule).

Two groups of chemists led by Evans⁶ and Nicolaou¹ independently achieved these Herculean tasks, producing a vancomycin aglycone molecule that lacks two sugars, which are normally attached to the side chain of amino-acid 4 (Fig. 1). Their approaches were quite different. Evans and his group first produced an intermediate with the bond linking side chains of amino-acids 5 and 7 having the wrong geometry, and subsequently fixed it. In contrast, the Nicolaou group made this bond correctly at an early point in their procedure, and it was maintained as such throughout the rest of the synthesis. But, when Nicolaou and co-workers crosslinked the side chains of amino-acids 2 and 4, they obtained a mixture (containing products both with the chlorine pointing 'forward' and 'backward'), and this mixture had to be separated by chromatography. The corresponding step in the synthesis by the Evans group gave, with good selectivity, the required 'backward pointing' chlorine atom. Both syntheses represent major achievements, and demonstrate the power of modern synthetic organic chemistry. But it would be a surprise if total synthesis proved to be a viable route to more effective versions of vancomycin. The quantities of antibiotics that are required in clinical work are so large that low-yield multi-step syntheses are impractical.

The latest achievement by Nicolaou's group is the conversion of vancomycin aglycone to vancomycin by the addition of the two sugars¹. This is a significant piece of work because it demonstrates that the selective addition of modified sugars to van-

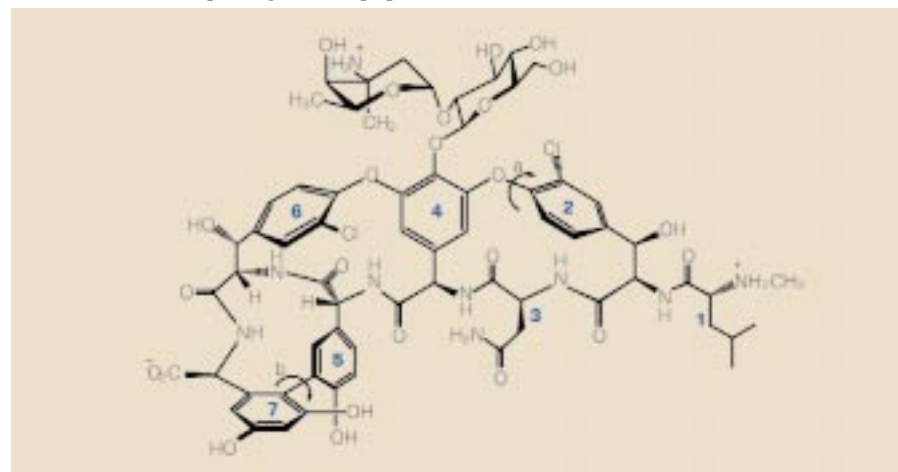


Figure 1 The structure of vancomycin with its constituent amino acids numbered 1 to 7. The two sugars, which have been added to synthetic vancomycin aglycone by Nicolaou *et al.*¹, are attached at the top of the benzene ring of amino acid number 4. Note how the chlorine atom (Cl) attached to the sidechain of amino acid 2 is directed away from the viewer, whereas that of the sidechain of amino acid 6 is directed towards the viewer. At room temperature, the benzene rings that carry these chlorine atoms cannot rotate through 180° (see arrow 'a'), and so reversal of the directions in which the chlorine atoms 'point' is precluded. Similarly, at room temperature, a 180° rotation about the bond linking the benzene rings numbered 5 and 7 cannot occur (see arrow 'b'), and so only the geometry shown is allowed in the final synthetic product.

comycin aglycone (which is readily available from vancomycin itself by acid hydrolysis) is possible in principle. It has been established that the sugars of the glycopeptides are important in promoting their activity⁸, and so vancomycins with modified sugars may well have enhanced activity against both VRE and MRSA. The viability of this proposal is suggested by the fact that Eli Lilly scientists have already directly modified these sugars, and in so doing have produced a glycopeptide antibiotic (now in trials) with greatly enhanced activity against VRE⁹. □
Dudley H. Williams is in the Department of Chemistry, Lensfield Road, University of

Cambridge, Cambridge CB2 1EW, UK.
e-mail: dhw1@cam.ac.uk

1. Nicolaou, K. C. *et al.* *Angew. Chem. Int. Edn* **38**, 240–244 (1999).
2. Neu, H. C. *Science* **257**, 1064–1073 (1992).
3. Williams, D. H. *Nat. Prod. Rep.* 469–477 (1996).
4. Arthur, M., Molinas, C., Depardieu, F. & Courvalin, P. *J. Bacteriol.* **175**, 117–127 (1993).
5. Walsh, C. T., Fisher, S. L., Park, I.-S., Prahalad, M. & Wu, Z. *Chem. Biol.* **3**, 21–28 (1996).
6. Evans, D. A. *et al.* *Angew. Chem. Int. Edn* **19**, 2700–2704; 2704–2708 (1998).
7. Nicolaou, K. C. *et al.* *Angew. Chem. Int. Edn* **19**, 2708–2714; 2714–2716; 2717–2719 (1998).
8. Gerhard, U., Mackay, J. P., Maplestone, R. A. & Williams, D. H. *J. Am. Chem. Soc.* **115**, 232–237 (1993).
9. Nicas, T. I. *et al.* *Antimicrob. Agents Chemother.* **49**, 2194–2199 (1996).

Gene expression

The amazing demethylase

Howard Cedar and Gregory L. Verdine

DNA methylation is important for mediating repression of gene expression during mammalian development. But although much is known about how methyl groups are added to DNA, little has been done to work out how they are taken off. In an exciting article on page 579 of this issue, Bhattacharya *et al.*¹ report that they have cloned a hitherto unknown gene which codes for a protein with a remarkable property — it can catalyse demethylation by directly removing methyl groups from 5-methyl-cytosine residues in DNA.

During normal development, cytosines in the dinucleotide CpG undergo a sequence of methylation and demethylation events, which are programmed to set up a defined methylation pattern (in both somatic and germline cells). This, in turn, plays a role in regulating gene expression. Soon after fertilization, a massive wave of demethylation erases almost all methyl groups in the genome, but the basic pattern is re-es-

tablished at about the time the embryo is implanted in the uterus^{2,3}. This remethylation is mostly nonspecific, but CpG islands (mainly found in the promoter regions of 'housekeeping' genes) are protected from this process — probably by specific demethylation⁴. Later in development, many tissue-specific genes are demethylated in the cell type in which they are expressed. In all of these cases, the conversion of 5-methyl-cytosine to cytosine takes place through an active process which does not require DNA replication^{4–6}.

But how does this process occur? To find out, Bhattacharya *et al.*¹ assumed that any enzyme which catalyses demethylation should be able to recognize methylated CpG dinucleotides in the DNA. So, they searched a database for DNA sequences homologous to the methyl-CpG-binding domain of a protein that is already known to bind CpG islands⁷. Remarkably, when the authors synthesized protein from one of these comple-

mentary DNAs, it had clear-cut demethylase activity *in vitro* (as determined by its ability to make a methylated plasmid sensitive to *HpaII* digestion). The authors also showed, by chromatography, that ³²P-labelled 5-methyl-cytosine nucleotides in DNA are converted to the cytosine form (Fig. 1), and that this is done without affecting the DNA backbone. Furthermore, this newly formed base must be cytosine, because it can then serve as a substrate for enzymatic remethylation *in vitro*. This experiment leaves open the possibility that the reaction involves the removal and then replacement of the cytosine base, but it is very unlikely that this could be accomplished by a single purified protein, without additional substrates or other enzymatic activities. So, the data indicate that the cytosine product must be generated by direct removal of the methyl group.

Although Bhattacharya *et al.* have not yet fully characterized all of the substrates and products of this reaction, they cite preliminary data indicating that 5-methyl-cytosine may react with water to produce cytosine and methanol. A quick calculation based on the kind of bonds broken (one O–H, one C–C) and formed (one C–O, one C–H) in such a reaction (Fig. 1) suggests that demethylation by water should be thermodynamically accessible, especially because the concentration of water is much higher than that of the methylated DNA.

Is there a plausible kinetic pathway that could lower the activation barrier enough for the reaction to take place? Here, also, the answer is probably yes. The main mechanistic problem in both methylation and demethylation has to do with the fact that C5 is a vinyl carbon. The energy of any intermediate with localized negative charge on this carbon would be unattainably high under physiological conditions. DNA methyltransferases (which methylate cytosine in DNA) solve this problem using a combination of covalent catalysis at C6 and proton shuffling at N3 (ref. 8). Given the ability of DNA methyltransferase to stabilize the intermediate, there should be no kinetic problem with a hydroxide ion attacking the C5 methyl group in the reverse reaction (Fig. 1). Simply put, the demethylase itself could turn the pyrimidine ring into a good leaving group. Tantalizingly, the demethylase even contains a tripeptide sequence (PLC) that may be similar to a sequence (PC or PPC) found at the active catalytic site of all cytosine methyltransferases.

Bhattacharya *et al.* are not the only ones to have isolated an enzymatic activity for carrying out demethylation. Jost's laboratory⁹ has characterized a highly purified glycosylase, which removes the 5-methyl-cytosine base in a reaction that uses RNA to recognize mCpG residues in the DNA. This reaction could potentially be the initial step in a complex demethylation mechanism, in

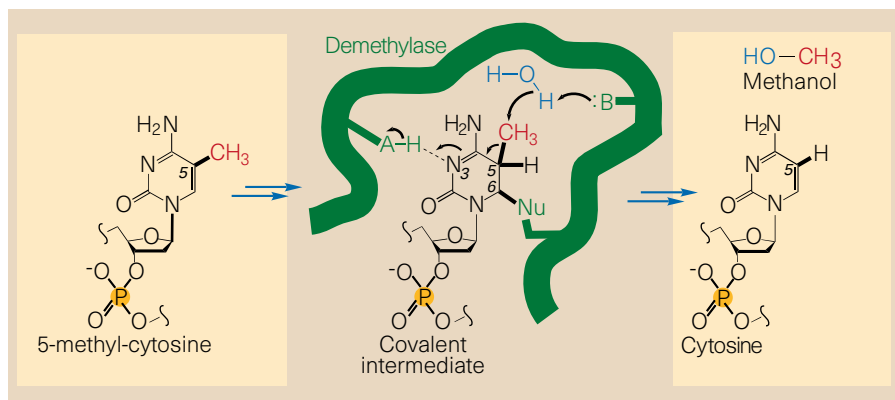


Figure 1 Mechanism for the enzymatic demethylation of 5-methyl-cytosine. The demethylase (green) is envisaged to form a covalent intermediate by addition of an enzymatic nucleophile (Nu–H) across the 5,6 double bond, assisted by proton shuffling at N3. This intermediate is poised to attack the hydroxide ion, which is generated by *in situ* activation of water. Double arrows indicate two reaction steps, with the intermediates not shown. In the case of enzymatic methylation, an analogous covalent intermediate is formed, but is further processed by cleavage of the C5–H bond as opposed to the C5–CH₃ bond. The 3'-phosphate labelled with ³²P in the tracer studies of Bhattacharya *et al.*¹ is in yellow.

which apyrimidinic acid residues would then be converted to cytosine *in vivo*. But, because this protein co-purifies with a G–T mismatch glycosylase, it is more likely to be involved in DNA repair, the reaction with 5-methyl-cytosine being simply fortuitous.

Weiss *et al.*¹⁰ have also detected an activity in crude cell extracts, which seems to carry out demethylation by some form of dinucleotide-exchange reaction. This activity can do cell-type- and gene-specific demethylation *in vitro*, suggesting that it could catalyse some of the demethylation events that occur during development. But this enzymatic system is still a long way from being fully characterized. For example, it was originally reported that this demethylation may involve RNA, but more recent results¹¹ indicate that partially purified preparations are not sensitive to RNase treatment.

It is now important to figure out what the new enzyme might do *in vivo*. Bhattacharya *et al.*⁴ have already shown that the demethylase gene is transcribed in a number of different normal tissues, but the corresponding demethylase activity was purified from only one cancer cell line. This is worth noting, because many tumours are characterized by an overall decrease in the levels of genomic methylation¹², and this may be an important mechanism for altering gene

expression and genome stability¹³ in the transformed state. □

Howard Cedar is in the Department of Cellular Biochemistry, Hebrew University Medical School, Jerusalem 91120, Israel.

e-mail: cedar@md2.huji.ac.il

Gregory L. Verdine is in the Department of Chemistry and Chemical Biology, and the Institute of Chemistry and Cell Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA.

e-mail: verdine@chemistry.harvard.edu

1. Bhattacharya, S. K., Ranchandi, S., Cervoni, N. & Szyf, M. *Nature* **397**, 579–583 (1999).
2. Monk, M., Boubelik, M. & Lehnert, S. *Development* **99**, 371–382 (1987).
3. Kafri, T. *et al.* *Genes Dev.* **6**, 705–714 (1992).
4. Frank, D. *et al.* *Nature* **351**, 239–241 (1991).
5. Paroush, Z., Keshet, I., Yisraeli, J. & Cedar, H. *Cell* **63**, 1229–1237 (1990).
6. Kafri, T., Gao, X. & Razin, A. *Proc. Natl Acad. Sci. USA* **90**, 10558–10562 (1993).
7. Hendrich, B. & Bird, A. *Mol. Cell. Biol.* **18**, 6538–6547 (1998).
8. Verdine, G. L. *Cell* **76**, 197–200 (1994).
9. Jost, J.-P., Fremont, M., Siegmund, M. & Hofsteenge, J. *Nucleic Acids Res.* **25**, 4545–4550 (1997).
10. Weiss, A., Keshet, I., Razin, A. & Cedar, H. *Cell* **86**, 709–718 (1996).
11. Swisher, J. F. A., Rand, E., Cedar, H. & Pyle, A. M. *Nucleic Acids Res.* **26**, 5573–5580 (1998).
12. Laird, P. W. & Jaenisch, R. *Annu. Rev. Genet.* **30**, 441–464 (1996).
13. Chen, R. Z., Pettersson, U., Beard, C., Jackson-Grusby, L. & Jaenisch, R. *Nature* **395**, 89–93 (1998).

rise to both neuronal and glial progeny.

It now comes as a surprise that these neural stem cells may reside not in the subventricular zone but, according to Johansson *et al.*¹, in the overlying ventricular lining. This layer, the ependyma, has all the features of a differentiated, post-mitotic epithelium, and the authors provide compelling evidence that the ependymal cells are neural stem cells. These cells were found to express immature neural markers, consistent with a stem-cell function. In addition, they normally divide very rarely. Johansson *et al.* found that, on average, only two cells in the entire mouse brain were labelled after a retroviral vector (which incorporates only into dividing cells) was injected into the lateral ventricle.

Johansson *et al.* propose that the ependymal stem cell divides asymmetrically to form one daughter cell, which stays as an undifferentiated stem cell in the ependymal layer, and another cell that moves into the subventricular layer below. Here it gives rise to a rapidly dividing progenitor cell pool, which provides a source for the neuronal and glial precursors that migrate towards their final destinations (Fig. 1). The authors found that after injury to the rat spinal cord, division of the ependymal stem cells increased dramatically to generate migratory astrocytes within the damaged area. Johansson *et al.* point out that cells with similar properties have also been shown in deeper parts of the brain⁶, but it is not known whether they are an additional pool of stem cells, distinct from the ependymal cells.

Neural stem cells can be stimulated to divide *in vitro* by epidermal growth factor or fibroblast growth factor-2. If these cells are transplanted into the intact brain they integrate and differentiate into a range of neuronal and glial cells^{6,7}. Yet it is difficult to show conclusively that they fulfil one of the classic requirements for a stem cell — the ability to regenerate a complete tissue (that

Stem cells

Breaking the brain–blood barrier

Anders Bjorklund and Clive Svendsen

As we develop, our cells become increasingly more specialized to carry out particular functions. Nonetheless, many cells retain a full complement of DNA and, in some cases, the capacity to regenerate a whole new animal. Take Dolly the sheep, for example. Although, in this case, DNA was transferred into a primitive egg cell, the fact that this DNA could be reprogrammed to encode an entire organism raised a fundamental question — how 'fixed' is any cell within an organism? Reporting in *Cell*, Johansson *et al.*¹ now show that cells within the ependymal lining of the adult brain ventricles may be multipotent neural stem cells that can generate new neurons and glia. And in *Science*, Bjornson *et al.*² describe how similar cells can regenerate blood tissues when transplanted into an irradiated mouse.

We are used to associating stem cells with renewable tissues such as blood, gut and skin. But some cells in the adult central nervous system have the capacity to generate new neurons and glial cells (astrocytes and oligodendrocytes) and, as such, they are considered to be neural stem cells^{3–5}. These remarkable cells can be isolated from the subventricular zone in the wall of the lateral

ventricle of the brain, and they divide in response to epidermal growth factor and fibroblast growth factor-2. But their exact nature has proved elusive. They have been proposed⁴ to reside in the subventricular zone, where they constitute a small population (0.1–1%) of relatively quiescent cells. On dividing (which is a rare event), they give

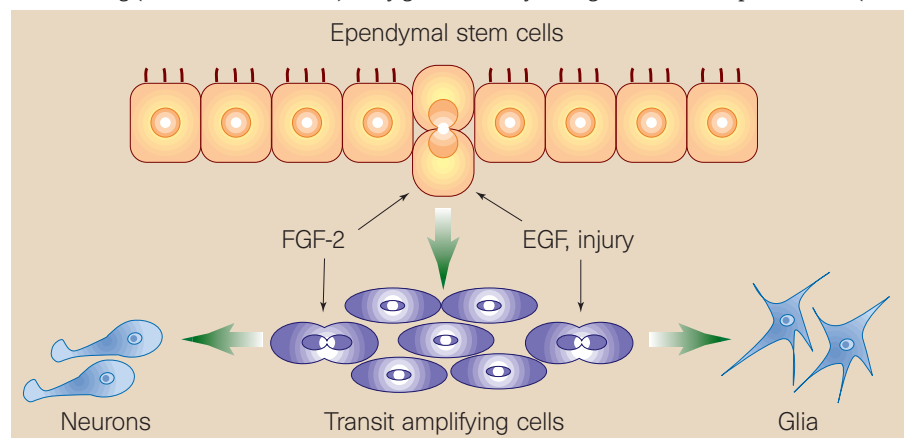


Figure 1 Generation of neural stem cells, based on the findings of Johansson *et al.*¹. The ependymal cells divide very rarely to generate a rapidly proliferating progenitor, which migrates to the underlying subventricular zone and can form both neurons and glia. Fibroblast growth factor-2 (FGF-2) and epidermal growth factor (EGF), as well as tissue injury, stimulate cell division in both stem cells and their multipotent progeny. Additional stem cells, yet to be identified, may reside in this subventricular pool.

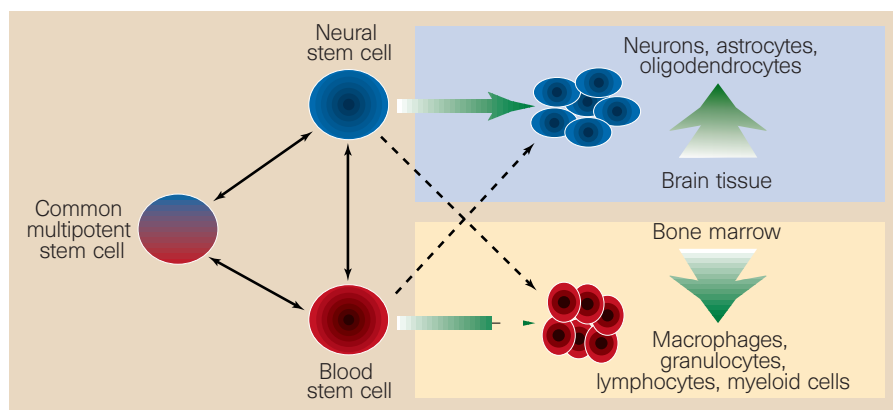


Figure 2 Possible ways in which neural and blood stem cells could interact. Bjornson *et al.*² have found that neural stem cells may differentiate into blood tissues under the influence of environmental signals from the bone marrow (dotted line). Alternatively, they may transform into a blood stem cell, or even de-differentiate into a more primitive neural/blood stem cell first (double-headed arrows). Contact with the bone marrow is important because neural stem cells differentiated *in vitro* can only make neural tissue (even when given growth factors that induce differentiation of blood cells). Although not as well characterized, blood stem cells may also give rise to neural cells (astrocytes) when exposed to the brain environment.

is, a full complement of cells with the appropriate architecture) after injury⁸. For this reason, their stem-cell status has been questioned⁴.

One system in which stem cells are well established is the blood. Here, haematopoietic stem cells can regenerate all cellular components in irradiated animals⁸. Although the relationship between brain and blood is complex, it has long been recognized that the brain contains a phagocytic cell type called microglia, which arise either from the neuroepithelium or from the bone marrow⁹. During the life of an animal, many cells from the adult bone marrow may enter the brain and, although most of them make microglia, a small percentage may develop into astrocytes¹⁰.

Now there is evidence of travel in the opposite direction. Bjornson *et al.*² generated neural stem cells from adult ROSA26 mice (which carry a specific cell marker), and isolated clones that gave rise to neurons, astrocytes and oligodendrocytes. Importantly, no blood cells were seen to arise from these cells *in vitro*. The authors then injected the clonal cultures into the tail veins of irradiated mice from another strain. Amazingly, 20 weeks later, both the blood and bone marrow from the grafted mice contained a range of blood cells derived from the transplanted neural stem-cell pool. How were these cells from the brain transformed into blood cells? The precursors from the central nervous system must have been able to respond to environmental cues in the irradiated marrow — they may either have reverted to a pluripotent phenotype, or have been induced to form blood cells directly (Fig. 2).

The findings of Johansson *et al.*¹ and Bjornson *et al.*² indicate that neural stem cells are much more plastic, and more widely spread throughout the adult central nervous

system, than previously thought. The exact biological function of these cells remains to be clarified, but the old dogma that the adult brain cannot produce new neurons can now safely be rejected. Most of the neural stem cells lie dormant, waiting for an activating stimulus. Yet their ability to generate neurons and glia, and their presence in the central nervous system throughout life, suggests new, intriguing possibilities for recovery and repair after damage to the central nervous system — and, unexpectedly, the regeneration of blood tissues. We do not know whether human neural stem cells also arise from the ependymal layer, or whether they have the capacity to turn into blood. However, similar embryonic human cells can be cloned¹¹, grown for extended periods¹² and continue to reside in the adult brain¹³, so it may not be long before we find out. □

Anders Bjorklund is at the Wallenberg Neuroscience Center, Lund University, Solvegatan 17, S-22362, Sweden.

e-mail: anders.bjorklund@mphy.lu.se

Clive Svendsen is at the MRC Centre for Brain Repair, University of Cambridge, Forvie Site, Robinson Way, Cambridge CB2 2PY, UK.

e-mail: cns1000@hermes.cam.ac.uk

- Johansson, C. B. *et al.* *Cell* **96**, 25–34 (1999).
- Bjornson, C. R., Rietze, R. L., Reynolds, B. A., Magli, M. C. & Vescovi, A. L. *Science* **283**, 534–537 (1999).
- McKay, R. D. G. *Science* **276**, 66–71 (1997).
- Weiss, S. & van der Kooy, D. *J. Neurobiol.* **36**, 307–314 (1998).
- Alvarez-Buylla, A. & Temple, S. J. *Neurobiol.* **36**, 105–110 (1998).
- Gage, F. H., Kempermann, G., Palmer, T. D., Peterson, D. A. & Ray, J. J. *Neurobiol.* **36**, 249–266 (1998).
- Takahashi, M., Palmer, T. D., Takahashi, J. & Gage, F. H. *Mol. Cell Neurosci.* **12**, 340–348 (1998).
- Potten, C. S. & Loeffler, M. *Development* **110**, 1001–1020 (1990).
- Thee, D. P. & Streit, W. J. *Glia* **7**, 5–8 (1993).
- Eglitis, M. A. & Mezey, E. *Proc. Natl Acad. Sci. USA* **94**, 4080–4085 (1997).
- Flax, J. D. *et al.* *Nature Biotechnol.* **16**, 1033–1039 (1998).
- Svendsen, C. N. *et al.* *J. Neurosci. Methods* **85**, 141–152 (1998).
- Eriksson, P. S. *et al.* *Nature Med.* **4**, 1313–1317 (1998).

Daedalus

Space aeronautics

Any spacecraft has to propel itself by firing something in the opposite direction — combustion gases for a conventional rocket, a plasma stream for an ion engine, reflected sunlight for a solar sail. Daedalus likes the idea of a motor which acts on the tenuous gas in space itself. A conventional propeller, of course, could never be made big enough. But Daedalus recalls the Crookes radiometer, the little 'light mill' in a glass bulb, which spins in sunlight. The gas in its bulb has to be so rarefied that the mean free path exceeds the thickness of its vanes, making it an ideal engine for a low-pressure environment. In such attenuated conditions, a temperature difference across a vane, set up by sunlight shining on one side of it, drives the surrounding gas past it. Each molecule that hits the edge of a vane recoils with increased thermal energy, propelling the vane.

Even at normal orbital altitudes the mean free path is many metres. So any large object with a temperature difference from front to back should be propelled through space, cold end first, by the radiometer effect. The obvious design is a modified solar sail. Daedalus proposes a long plastic sheet, perhaps many square kilometres in area but only a few micrometres thick, with a surface coating graded continuously from dead black at the 'rear' to fully metallic at the 'front'. The blacker the surface, the more it would be heated by sunlight. A temperature gradient would be established along the length of the sheet, propelling it cold-end first by thrust from its whole vast area. Radiation-pressure would also act on the sheet, making it a combined radiometer craft and solar sail. Unlike a pure solar sail, it could be oriented (by a system of giant adjustable veins or rudders) to steer towards the Sun, the ideal direction for a solar-powered spacecraft.

For journeys away from the Sun, Daedalus will impregnate his radiometer craft with a graded distribution of radioactive material, to keep the rear always hotter than the front. As it slowly climbs away from the Sun in an outward spiral orbit, its solar propulsion will fade but its radioactive drive will keep going. In a mission lasting centuries, it could journey far into interstellar space. Its transmitters, powered by thermoelectric radioactive heating and aimed at the Earth by a huge parabolic section of its metallized film construction, would keep us informed of its findings.

David Jones

Superalliance of bottlenose dolphins

It is quite common to find several levels of nested male alliances in human political organization^{1,2} but these are extremely rare in other species³. Yet we found that male bottlenose dolphins (*Tursiops* sp.) at Shark Bay, Western Australia, form two levels of alliance within a social network of more than 400 individuals. Fourteen of the males formed highly labile alliances, rather than the more typical stable ones, and joined forces in a large 'superalliance' that competed directly with smaller teams of stable alliances.

Alliances of two or three male bottlenose dolphins have been reported in Shark Bay and Sarasota Bay, Florida³⁻⁷. Alliances in both populations are strong and stable over long periods, lasting up to 12 years in Shark Bay and 20 years in Sarasota⁷. As a measure of the strength of the association between individuals, we calculated association coefficients from observations of joint participation in consorting a female⁵. Values range from zero for two individuals who were never allies, to 100 for two individuals who were always in the same alliance. Males in the same alliance³⁻⁸ typically had values of 80 to 100 (Fig. 1a), similar to those for mothers and their dependent offspring⁸.

Teams of two stable alliances were sometimes found to cooperate as 'second-order' alliances to attack other alliances or to defend against attacks. Stable alliances may maintain second-order alliance relationships with one or two other stable alliances, but only cooperate with one alliance at a time^{3,4}. Second-order alliance relationships do not generally endure for more than a few years, and alliances that normally cooperate may oppose each other in some social contexts³.

There were no stable alliances in the superalliance (Fig. 1b). The composition of alliances in the superalliance was highly labile, as males often switched partners (Table 1). Over three years, we found evidence of 39 different alliances among the 14 superalliance males, including 35 trios and 4 pairs. Of the 14 males, 9 had eight or more alliance partners, none of which was with males from outside the superalliance.

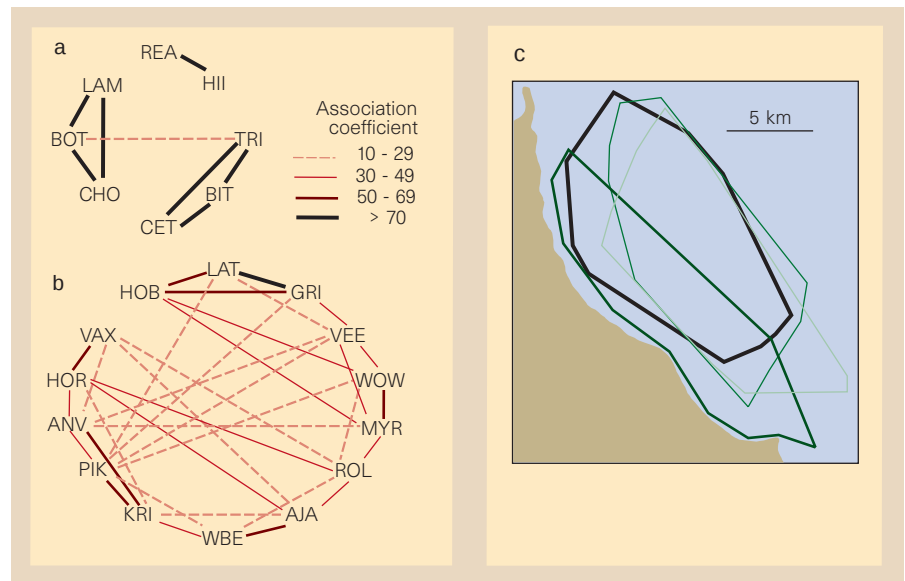


Figure 1 Contrasting patterns of alliance formation cannot be explained by habitat differences. a, b, 'Sociograms' showing the strength of association between: a, males in three stable alliances who formed second-order alliances together³; and b, males in the 14-member superalliance. The strength of the alliance is indicated by the association coefficient. c, The home range of the superalliance (black line) overlaps extensively with the ranges of three stable alliances. Each of the three stable-alliance ranges includes data from two stable alliances (5–6 individuals) that associated as a second-order alliance.

This lability was also evident within years: during June to October 1996, each of the 14 males was in 3–7 alliances and had 4–8 of the 14 males as alliance partners.

Analysis of association coefficients, group size (Table 1) and conflicts indicates that the superalliance is a very large second-order alliance. Association coefficients linking individuals in the superalliance (mean, 58; range, 34–97) were within the range linking stable male alliances that formed second-order alliances⁸.

We observed four conflicts involving between 6 and 14 members of the superalliance and teams of two stable alliances or a group of seven males who matured just before or early in the study. Superalliance males travelled up to 3 kilometres to join conflicts involving members of their group, and were victorious in each case.

The differences between the superalliance and stable alliances are not related

to age. Strong bonds are often evident before maturity in males who form stable alliances and may crystallize at maturity (age 10 to 12 years)^{6,7}. Minimum ages of 14 to 19 years are known for ten of the members of the superalliance, judging from the first time they were photographed. The benefits of large group formation may increase in habitats with high predation risk and the costs may be reduced in habitats with more resources. However, these possibilities cannot explain the formation of superalliances because of the extensive overlap in home range between stable alliances and the superalliance (Fig. 1c).

This variation in alliance formation within one social group has not to our knowledge been described for any other species. The closest parallel is in male baboons (*Papio cynocephalus*), in which alliance formation is a conditional strategy used mostly by mid-ranking males against high-ranking males⁹. We cannot exclude the possibility that the large size of the dolphin superalliance allows individuals to compete with teams of stable alliances.

The extreme lability of alliances in the superalliance is puzzling, but may provide a mechanism by which males in the superalliance maintain affiliative bonds. Large brain size has been linked to complex social relationships, and alliance formation in particular¹⁰⁻¹². Our findings suggest that the large brain size of bottlenose dolphins — only humans have larger brains relative to

Table 1 Males in stable alliances compared with males in the superalliance

	Stable alliances	Superalliance
Number of consortships	16–24	10–30
Number of alliances	1–2	5–11
Number of alliance partners	1–3	5–11
Alliance partners with primary alliance (%)	84–100	17–57
Mean male group size	3.6	6.1

The numbers of consortships, different alliances (pairs and trios) and alliance partners, and the percentage of consortships by each male's most common (primary) alliance, are compared for males in stable alliances (8 males, 62 consortships) and males in the superalliance (14 males, 100 consortships). The mean male group size was larger for members of the superalliance ($n=57$) than for five stable alliances ($n=115$; Mann-Whitney U test, $P<0.001$). Stable alliances were never documented in male groups of 10 or more individuals, but 25% of groups with at least one member of the superalliance contained 10 or more of the 14 males.

body size — may be matched by the complexity of their social relationships.

Richard C. Connor*, **Michael R. Heithaus†**,
Lynne M. Barre†

*Biology Department,
University of Massachusetts, Dartmouth,
North Dartmouth, Massachusetts 02747, USA
e-mail: rconnor@umassd.edu

†Biology Department, Simon Fraser University,
Burnaby, British Columbia, V5A 1S6, Canada

‡Biology Department, Georgetown University,
Washington, DC 20057, USA

- Boehm, C. in *Coalitions and Alliances in Humans and Other Animals* (eds Harcourt, A. H. & de Waal, F. B. M.) 137–173 (Oxford Univ. Press, 1992).
- Falger, V. S. E. in *Coalitions and Alliances in Humans and Other Animals* (eds Harcourt, A. H. & de Waal, F. B. M.) 323–348 (Oxford Univ. Press, 1992).
- Connor, R. C., Smolker, R. A. & Richards, A. F. in *Coalitions and Alliances in Humans and Other Animals* (eds Harcourt, A. H. & de Waal, F. B. M.) 415–443 (Oxford Univ. Press, 1992).
- Connor, R. C., Smolker, R. A. & Richards, A. F. *Proc. Natl Acad. Sci. USA* **89**, 987–990 (1992).
- Connor, R. C. *et al. Behaviour* **133**, 37–69 (1996).
- Wells, R. S. in *Dolphin Societies* (eds Pryor, K. & Norris, K. S.) 199–225 (Univ. California Press, Berkeley, 1991).
- Connor, R. C. *et al. in Cetacean Societies: Field Studies of Whales and Dolphins* (eds Mann, J., Connor, R. C., Tyack, P. L. & Whitehead, H.) (Chicago Univ. Press, in the press).
- Smolker, R. A. *et al. Behaviour* **123**, 38–69 (1992).
- Nöe, R. *Anim. Behav.* **47**, 211–213 (1994).
- Harcourt, A. H. in *Coalitions and Alliances in Humans and Other Animals* (eds Harcourt, A. H. & de Waal, F. B. M.) 445–471 (Oxford Univ. Press, 1992).
- Cords, M. in *Machiavellian Intelligence* Vol. II, *Extensions and Evaluations* (eds Whiten, A. & Byrne, R. W.) 24–49 (Cambridge Univ. Press, 1997).
- Barton, R. A. & Dunbar, R. I. M. in *Machiavellian Intelligence* Vol. II, *Extensions and Evaluations* (eds Whiten, A. & Byrne, R. W.) 240–263 (Cambridge Univ. Press, 1997).

Anterior sphenoid in modern humans

Lieberman has proposed¹ that reduced midfacial projection (MFP), in which most of the face lies beneath the neurocranium, is a major unique, derived character of anatomically modern *Homo sapiens*, and that this reduction is largely a consequence of reduced anterior sphenoid length (ASL). Lieberman's conclusions were based on comparisons of a small sample of archaic *Homo* crania with those of Holocene and Pleistocene anatomically modern *H. sapiens*. We have made new measurements of ASL and MFP, and find that ASL was incorrectly estimated in those archaic fossil crania in which these landmarks are unambiguously preserved. It turns out that the anterior sphenoid in modern humans is no shorter than in archaic *Homo*.

The new measurements were taken from better-quality radiographs and computed tomography scans^{2,3} and from the original specimens of Gibraltar 1 and Broken Hill (courtesy of C. Stringer, T. Molleson and F. Zonneveld). ASL values in Holocene and Pleistocene modern humans are 19.9 mm (s.d. 2.0) and 20.0 mm (s.d. 1.8), respectively¹,

not significantly different ($P > 0.05$, Scheffé's F) from those of archaic *Homo* (Gibraltar 1, 17.2 mm; Monte Circeo, 16.9 mm; Broken Hill, 17.2 mm). Figure 1a confirms that reduced MFP in anatomically modern humans is not associated with a shorter ASL.

To assess the spatial relationships of ASL and MFP in relative terms, we did a geometric morphometric analysis comparing Holocene modern human crania with the three archaic *Homo* fossils (Fig. 1b,c). The transformation grid indicates that, relative to the size of the landmark configuration, MFP is shortened and ASL is lengthened in Holocene *H. sapiens*. The factors underlying these changes may include facial reduction, increased basicranial flexion, and expansion of the temporal lobes in the middle cranial fossae. The comparison also suggests that the pharyngeal area between the palate and the foramen magnum is anteroposteriorly constricted in Holocene

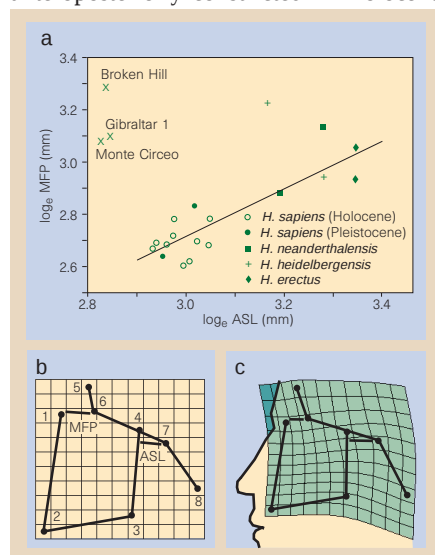


Figure 1 Analysis of measurements. **a**, Plot of ASL and MFP (after Fig. 3a of ref. 1), with new measurements of Broken Hill, Gibraltar 1 and Monte Circeo. MFP is the distance from nasion to foramen caecum perpendicular to the posterior maxillary plane. ASL is the minimum distance from sella to the posterior maxillary plane. **b**, **c**, Geometric morphometric comparison of eight facial, neurocranial and basicranial landmarks from 28 combined-sex Holocene *H. sapiens* skulls (Indian subcontinent), Monte Circeo, Gibraltar 1 and Broken Hill. Principal components analysis of tangent coordinates following generalized Procrustes analysis yields 12 non-zero eigenvectors⁴. Only principal component I (PC I; 37% total shape variance) separates fossil from extant specimens. The shapes represent means for the fossils (**b**) and Holocene *H. sapiens* (**c**), both on PC I and each rescaled to their respective mean size. A cartesian transformation grid (thin plate spline⁵) from fossil to extant means is superimposed. Numbers: 1, nasion; 2, prosthion; 3, maxillary tuberosity; 4, laterally projected average intersection between greater wings and planum sphenoidale; 5, anteriormost point of cranial cavity; 6, foramen caecum; 7, sella; 8, basion (estimated in Neanderthals); 3 to 4, posterior maxillary plane.

modern humans, as was inferred by Lieberman¹, but that this is unrelated to ASL.

We conclude that, although ASL is intraspecifically correlated with MFP in modern humans and chimpanzees¹, it does not account for the unique form of the modern human cranium. Our analysis highlights the need for research that integrates comparative morphometric analyses with developmental studies of cranial growth in human and non-human primates.

Fred Spoor*, **Paul O'Higgins***,
Christopher Dean*, **Daniel E. Lieberman†**

*Department of Anatomy and Developmental
Biology, University College London,
21 University Street, London WC1E 6JJ, UK
e-mail: f.spoor@ucl.ac.uk

†Department of Anthropology,
George Washington University,
Washington, DC 20052, USA
e-mail: danlieb@gwu.edu

- Lieberman, D. *Nature* **393**, 158–162 (1998).
- Sergi, S. in *The Circeo 1 Neanderthal Skull: Studies and Documentation* (eds Piperno, M. & Scichilone, G.) 23–173 (Istituto Poligrafico e Zecca dello Stato, Rome, 1991).
- Silipo, P. *et al. in The Circeo 1 Neanderthal Skull: Studies and Documentation* (eds Piperno, M. & Scichilone, G.) 513–538 (Istituto Poligrafico e Zecca dello Stato, Rome, 1991).
- Dryden, I. L. & Mardia, K. V. *Sankhya A* **55**, 460–480 (1993).
- Bookstein, F. L. *IEEE Trans. Patt. Anal. Mach. Intell.* **11**, 567–585 (1989).

Iodine oxide in the marine boundary layer

A striking example of the influence of halogen chemistry on tropospheric ozone levels is the episodic destruction of boundary-layer ozone during the Arctic sunrise by reactive halogen species^{1,2}. We detected iodine oxide in the boundary layer at Mace Head, Ireland (53°20' N, 9°54' W) during May 1997, which indicates that iodine chemistry is occurring in the troposphere.

Reactive halogen species in the atmosphere act as catalysts in several photochemical reaction cycles that are closely linked with ozone³. Iodine atoms react preferentially with ozone, forming iodine oxide, IO. IO can react with itself or with the halogen oxides BrO and ClO to produce O₂ and halogen atoms. If these react with ozone, a catalytic mechanism destroys two ozone molecules per cycle. The reaction of IO with HO₂ forms HOI, which is rapidly photolysed into I and OH. This catalytic cycle also effectively destroys ozone.

The measurement of IO in the boundary layer has so far been unsuccessful. The upper limit of the mixing ratios, determined as 0.5–0.9 part per trillion (p.p.t.)^{3,4}, agreed with model predictions^{5–9}. We measured the concentrations of O₃, NO₂, SO₂, HCHO, HONO, BrO, ClO and IO by using long-path differential optical absorption spectroscopy (LP-DOAS)¹⁰ during the

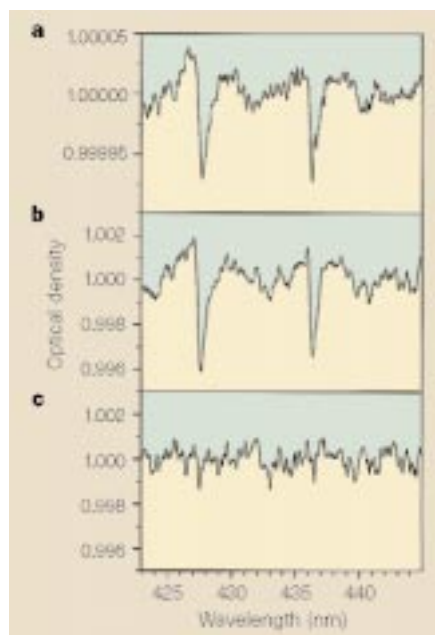


Figure 1 Spectra. **a**, IO reference spectrum measured in the laboratory. **b**, Atmospheric absorption spectrum measured with DOAS¹⁰ at a spectral resolution of 0.27 nm, after analysis. A fitting procedure scaled and subtracted the absorption of water (at 441–445 nm) and NO₂ which was below the detection limit. IO absorption bands are visible and agree in spectral position with the absorption cross-section¹² used to determine the mixing ratio of 6.6 ± 0.5 p.p.t. **c**, Residual spectral structure of the analysis, which determines the quality of the measurement.

HALOTROP/ACSOE field campaign at the atmospheric research station in Mace Head from 21 April to 30 May 1997. DOAS identifies and quantifies trace gases by their specific narrow-band (less than 5 nm) optical absorption structure in the open atmosphere, separating trace gas absorptions from broad-band molecule and aerosol extinction processes. A measured atmospheric spectrum after analysis is compared with a reference spectrum of IO in Fig. 1. The characteristic absorption bands of IO at 427.6 nm and 436.4 nm are both visible.

During 5–8 May 1997, IO concentrations increased above the detection limit of 0.9 p.p.t. The diurnal variation of IO follows solar radiation (Fig. 2). The ozone concentration during this period was about 30 ± 10 parts per billion (p.p.b.). The wind speed (Fig. 2) was higher than during the rest of the campaign. The wind was from the north, resulting in very clean air from the ocean. The correlation of IO concentrations with solar radiation indicates that iodine is produced by the photolysis of precursor species. Wind direction and preliminary trajectory analysis indicate that these species come from the ocean or the nearby shore. As there were increased concentrations of the short-lived iodinated hydrocarbons CH₂I₂ and CH₂BrI (ref. 11), they were the most likely precursors.

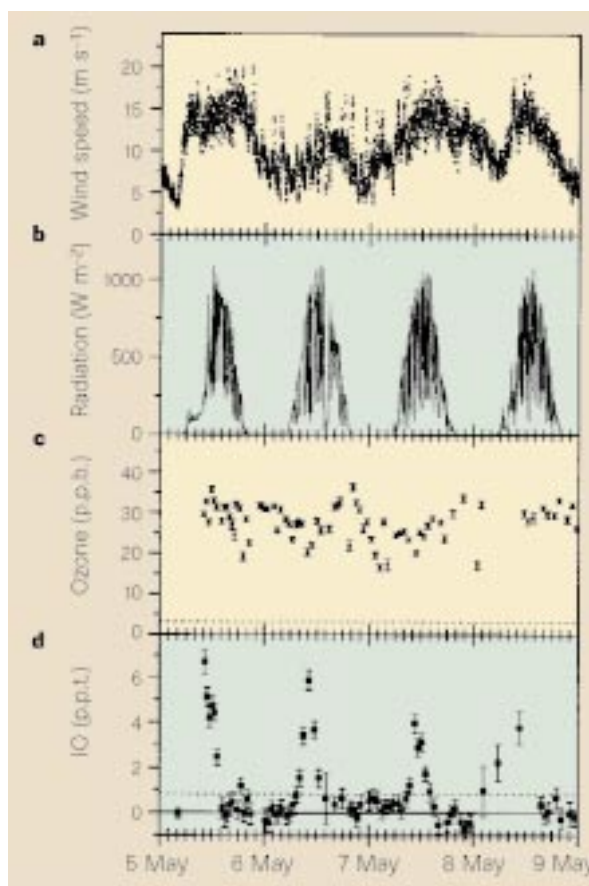


Figure 2 IO and ozone measurements at Mace Head. Iodine levels were influenced by: **a**, wind speed; and **b**, global radiation. Mixing ratios and 1σ errors of: **c**, ozone; and **d**, IO. Broken lines in **c** and **d** indicate the detection limit, defined as the average of twice the 1σ errors in this period. Trace gases were below their respective detection limits of 1.7 p.p.t. BrO, 34 p.p.t. ClO, 70 p.p.t. NO₂, 170 p.p.t. HONO, 0.8 p.p.b. SO₂ and 0.4 p.p.b. HCHO. In the rest of the campaign, ozone varied between 15 and 50 p.p.b. Maxima of 4 p.p.b. NO₂ and up to 6 p.p.b. SO₂ were observed in polluted air masses.

We used a photochemical box model (see Supplementary Information) to estimate the magnitude of the different ozone-loss processes for 6 p.p.t. IO at noon. A loss of 0.06 p.p.b. ozone per hour is calculated for the IO self-reaction cycle. Assuming an upper limit to the mixing ratio of 1 p.p.t. for BrO and ClO, the crossreactions with IO would destroy less than 0.03 and 0.01 p.p.b. per hour, respectively. From the model, we calculated that there was about 6 p.p.t. of HO₂, leading to an effective ozone destruction of 0.27 p.p.b. per hour by the HOI cycle. In the presence of IO, the modelled HO₂ concentration decreases from roughly 8 to 6 p.p.t. owing to the reaction HO₂ + IO. The ratio of [NO₂]/[NO] increases by about 30% as a result of the reaction NO + IO. At constant NO_x (comprising NO and NO₂) levels of 70 p.p.t., both effects slow down photolytic ozone production by 0.06 p.p.b. per hour. A larger reduction of O₃ production is expected for higher NO_x concentrations, up to a maximum of 0.15 p.p.b. per hour at 0.5 p.p.b. NO_x (see Supplementary Information).

The estimated total ozone loss of 0.39 p.p.b. per hour is less than the observed variability of the O₃ concentrations of about 10 p.p.b. but has to be compared with the removal rate by dry deposition of around 0.1 p.p.b. per hour over the ocean and 1 p.p.b. per hour over land. Gas-phase reactions, such as photolysis to form OH radicals, destroy about 0.2

p.p.b. per hour of O₃ at noon. A mixing ratio of 6 p.p.t. IO can therefore increase the removal rate of ozone by about 70% over the ocean and 10% over land.

Our measurements show that IO, although recorded only over four days, exceeds the predicted tropospheric concentrations^{5–9}. At levels of 6 p.p.t., IO can have a substantial influence on boundary-layer photochemistry. From our results and other data¹¹, we predict that IO should be present in clean air on coasts populated by macroalgae. Further investigation should establish the conditions that give rise to IO and the frequency of concentration increases.

Björn Alicke, Kai Hebestreit, Jochen Stutz, Ulrich Platt

Institut für Umweltphysik, University of Heidelberg, INF 366, D-69120 Heidelberg, Germany
e-mail: stu@uphys1.uphys.uni-heidelberg.de

1. Bottenheim, J. W. *et al.* *J. Geophys. Res.* **95**, 18555–18568 (1990).
2. Hausmann, M. & Platt, U. *J. Geophys. Res.* **99**, 399–413 (1994).
3. Wayne, R. P. *et al.* *Atmos. Environ.* **29**, 2675–2884 (1995).
4. Tuckermann, M. *et al.* *Tellus* **49B**, 533–555 (1997).
5. Chameides, W. L. & Davis, D. D. *J. Geophys. Res.* **85**, 7383–7398 (1980).
6. Jenkin, M. E. *AEA Environment and Energy Report AEA-EE-0405*, (AEA, 1992).
7. Solomon, S., Garcia, R. R. & Ravishankara, A. R. *J. Geophys. Res.* **99**, 20491–20499 (1994).
8. Sander, R. & Crutzen, P. J. *J. Geophys. Res.* **101**, 9121–9138 (1996).
9. Davis, D. *et al.* *J. Geophys. Res.* **101**, 2135–2147 (1996).
10. Platt, U. *Chem. Anal. Ser.* **127**, 27–83 (1994).
11. Carpenter, L. J. *et al.* *J. Geophys. Res.* (in the press).
12. Laszlo, B., Kurylo, M. J. & Huie, R. E. *J. Phys. Chem.* **99**, 11701–11707 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Horse sickness and ENSO in South Africa

African horse sickness (AHS) is the most lethal infectious horse disease, with mortality rates in susceptible animals of up to 95% (ref. 1). Since its discovery at Cape Colony in the early eighteenth century, major epizootics of this viral disease have occurred in South Africa once every 10 to 15 years on average; in the largest epizootic, more than 40% of the entire horse population died. However, the cause of these major epizootics has remained unknown until now. We have found a very strong association between the timing of these epizootics and the warm (El Niño) phase of the El Niño/Southern Oscillation (ENSO), and suggest that the association is mediated by the combination of rainfall and drought brought to South Africa by ENSO.

AHS is endemic to sub-Saharan Africa but epizootics are particularly frequent and severe in South Africa, which has a large population of susceptible horses, many zebra (the vertebrate reservoir of the AHS virus) and a widespread distribution of its principal insect vector, the biting midge *Culicoides imicola*. This vector and closely related species are widespread in the Old World. AHS outbreaks have occurred in Europe and Asia, with up to 300,000 equids dying in a single epizootic. Both North America and Australia, which have competent vectors, are also at risk.

ENSO, a periodic ocean-atmosphere fluctuation in the Pacific Ocean, is a major cause of climatic variability over much of the world, including southern Africa². ENSO events have been correlated with epidemics of several mosquito-borne diseases³⁻⁶ in parts of the world where ENSO teleconnections are strongest (the Pacific, South America and south Asia), although the associations themselves are not very strong and most data sets have been relatively short. In South Africa, the high value of horses and their extreme mortality rate from AHS means that good records were kept of AHS epizootics, allowing us to investigate historical patterns of the disease.

Since 1803, from which time there have been sufficient data to reconstruct El Niño events of at least moderate strength⁷, there have been 14 large epizootics of AHS in South Africa (1819, 1837 or 1839, 1854-55, 1862-63, 1877-78, 1887-88, 1891-92, 1913-14, 1918, 1923, 1925, 1940, 1953 and 1996). A list of 55 warm-phase ENSO events^{8,9} between 1803 and 1997 includes the following (part years omitted): 1819, 1837-39, 1854-55, 1862, 1877-78, 1888, 1891, 1914, 1918-19, 1923, 1925-26, 1940 and 1953. Of the 14 AHS epizootics, only

the most recent (1996) did not coincide with a warm-phase ENSO. We calculate the probability of an epizootic coinciding with an ENSO by chance to be 0.554 if ENSO events last two years. On this basis, we conservatively estimate the probability of our finding (13 of 14 epizootics coincided with ENSO) is at most 0.0032.

Warm-phase ENSOs bring both rainfall and drought to southern Africa¹⁰. Populations of the AHS virus vector, *C. imicola*, breed in damp soil and can increase 200-fold in years of heavy rain, raising the possibility that ENSO-brought rain may cause the association with major AHS epizootics. However, heavy rainfall occurs for other reasons in many non-ENSO years but epizootics do not result.

We investigated climatic patterns for the western region of South Africa, where most of the major AHS epizootics have occurred. We obtained monthly rainfall data (since records began in 1842) from the South Africa Weather Bureau and calculated the average rainfall in each three-month period

of the year. From these values, we estimated the rainfall anomaly (the percentage deviation from the mean rainfall in non-ENSO years) over the potential two-year time course of a warm-phase ENSO, from six months before the ENSO year to six months after it.

Warm-phase ENSOs typically have heavy rain followed by drought (Fig. 1a). Similar patterns are apparent in those warm-phase ENSO years when there was no epizootic of AHS (Fig. 1b). However, the average rainfall patterns are different in the warm-phase ENSO years when there were also epizootics of AHS (Fig. 1c). There is pronounced drought in the early part of the ENSO year, followed by heavier rain than usual between April and June. Heavy rainfall and drought are not limited to warm-phase ENSO events but the combination is favoured by them. Only 4 per cent of non-ENSO years combined drought from January to March and heavy rain from April to June, compared with nearly 20% of ENSO years (χ^2 , $P < 0.01$).

But how does a spell of drought followed by heavy rainfall affect AHS epizootics? Breeding sites of the insect vector may be altered or, during drought, the virus reservoir (zebra) may congregate near the few remaining sources of water where they are in contact with, and infect, more vectors. High temperature during droughts increases vector population growth rates and favours virus transmission. Although the reasons are not known, the significance of drought to AHS in South Africa has been long appreciated. In 1863, before the causative agent of the disease and its vector were known, or effects of ENSO on southern Africa were suspected, it was observed that "there are seasons (usually wet ones) succeeding drought, in which great mortality occurs among horses at the fall of the year"¹¹.

Matthew Baylis*†, Philip S. Mellor*, Rudy Meiswinkel†

*Institute for Animal Health, Ash Road, Pirbright, Surrey GU24 0NF, UK

e-mail: matthew.baylis@bbsrc.ac.uk

†Onderstepoort Veterinary Institute, Onderstepoort 0110, South Africa

‡Present address: Institute for Animal Health, Compton, Newbury, Berkshire RG20 7NN, UK

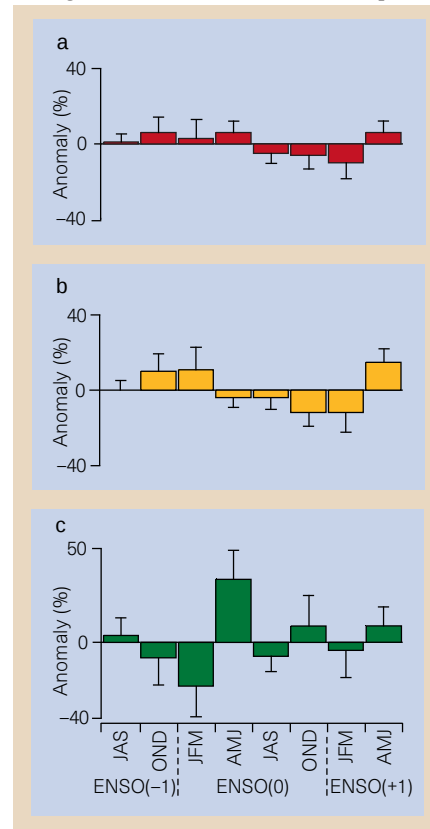


Figure 1 Average rainfall anomalies in western South Africa. a, Years of ENSO; b, years of ENSO without AHS epizootics; c, years of ENSO with AHS epizootics. The number of years averaged are 43, 32 and 11, respectively. Zero represents the average rainfall experienced in 113 non-ENSO years. Bars show percentage deviation (+s.e.) from this average for three-monthly intervals (JAS, July to September; OND, October to December; JFM, January to March; AMJ, April to June) over a two-year period. Data are means of rainfall anomalies at Cape Town Royal Observatory and Worcester.

- Hess, W. R. in *The Arboviruses: Epidemiology and Ecology* (ed. Monath, T.P.) 1-18 (CRC Press, Boca Raton, Florida, 1988).
- Nicholson, S. E. & Kim, J. *Int. J. Climatol.* **17**, 117-135 (1997).
- Nicholls, N. *Lancet* **342**, 1284-1285 (1993).
- Bouma, M. J. & van der Kaay, H. J. *Trop. Med. Int. Health* **1**, 86-96 (1996).
- Bouma, M. J. *et al. Trop. Med. Int. Health* **2**, 1122-1127 (1997).
- Hales, S., Weinstein, P. & Woodward, A. *Lancet* **348**, 1664-1665 (1996).
- Quinn, W. H., Neal, V. T. & de Mayolo, S. E. A. *J. Geophys. Res.* **92**, 14449-14461 (1987).
- Quinn, W. H. in *El Niño. Historical and Paleoclimatic Aspects of the Southern Oscillation* (eds Diaz, H. F. & Markgraf, V.) 119-149 (Cambridge Univ. Press, 1992).
- Webster, P. J. & Palmer, T. N. *Nature* **390**, 562-564 (1997).
- Ropelewski, C. F. & Halpert, M. S. *Mon. Weath. Rev.* **115**, 1606-1626 (1987).
- Paton, T. *Veterinarian* **36**, 489-494 (1863).

Unique memories: creating the mind's 'I'

Momentous Events, Vivid Memories
by David B. Pillemer
Harvard University Press: 1998. 212 pp. \$35,
£23.50

Martin Conway

We all remember, of that there can be no doubt. Whether we remember accurately or inaccurately, in detail or in abstract, are questions that researchers have investigated for many years. However, there is another, more fundamental question: why do we remember at all?

There are many animals on this planet that survive without having memories of critical moments in their lives, turning points, originating events. They have no vivid memories of sexual encounters, memories of loss, memories of love. Yet these animals appear to survive with an adaptive fitness equal to our own. So why is it that we have conscious recollection of the past?

David Pillemer's book is in part devoted to the question of why we have what memory researchers have come to call 'autobiographical memory' — memory for the events of our lives. Autobiographical memories are the things that make us what we are, the mental representations that distinguish you from me. Of course there are other types of knowledge that do this as well: one's personal beliefs, attitudes and so forth. Nevertheless, ultimately it is individual experience that separates one human being from another. As Pillemer demonstrates, two people who have the same experience, for example starting college together, will nonetheless have many different memories that reflect their personal projects and goals, whether conscious or not, during that period.

Pillemer devotes the middle portion of his book to a natural history of autobiographical memories. He assigns memories of different sorts of experiences to different classes, for example 'turning points', 'anchoring events' — events which retrospectively reveal a sequence of causality which one was not able to perceive at the time — and several other kinds of vivid memory.

Consider, for example, 'originating events'. This is a particularly interesting class of memory that Pillemer and his colleagues have investigated in some detail. It turns out that many people have memories of moments that originated life themes. Often these relate to experiences that occurred when the rememberer was at college or in high school, but sometimes the events date back to periods earlier in life. For instance, Einstein was shown a compass when he was four or five years old, and much later in life recalled the event. "That this needle behaved in such a determined way did not at all fit into



Each to his own: a group of Eton schoolboys experiencing a common event — in this case, a 1937 football match — that each one of them will remember differently.

the nature of events... I can still remember — or at least I believe I can remember — that this experience made a deep and lasting impression on me. Something deeply hidden had to be behind things." We have, perhaps, come to expect such musings from great men like Einstein. However, Pillemer's research shows that these are common experiences; many people have memories of originating moments when a particular stream of thought came into being — a stream of thought which was to persist, possibly, throughout life.

One of the interesting lines of thought in this book is a division of memory into images versus narratives. The image-based memory system is conceived as being phylogenetically older than the system based on narrative, which is conceived as being both more recent and wholly dependent on language.

It seems to me that a very sharp division along these lines is probably not going to capture the full nature of autobiographical memories. However, it is quite probable that there is an older emotion-based, image-based memory system mediated by older brain regions. For example, the temporal lobe memory system, including mid-brain structures such as the hippocampus and the amygdala, is older than more verbally based memory systems mediated by networks in the frontal lobes — a region of the brain which phylogenetically is comparatively recent. Indeed, one possibility is that these more recent memory networks act to modulate the output of the older memory systems, which is itself largely driven by cues or stimuli. The notion of characterizing these as

'image' versus 'narrative' memory systems captures important components of both types of system.

In the later part of the book the important role of culture and individual, as well as societal, meanings in autobiographical memory is given detailed coverage. The central point to emerge is that autobiographical memories are meaningful mental structures. As well as carrying personal meaning for the rememberer, they often carry a wider meaning for sub-groups within society and even for society as a whole. How we are to model and understand these meanings is as yet unknown, although undoubtedly any model will eventually have to draw not only upon the cognitive neurosciences, but also upon systems of thought developed in psychoanalytic traditions of thinking, which have for more than 100 years addressed issues relating to autobiographical memory.

Pillemer teases out these issues and they inevitably lead to a consideration of why we, as a species, have these rather curious mental representations. For Pillemer, part of the answer lies in his suggestion that autobiographical memories and the ability to have them provide a certain sort of social intelligence that could not be delivered in any other way. In this social intelligence, the specifics of experience are preserved and can be re-examined, thought over, reconceptualized and fed back into the group. I believe he is right in this.

I would only add that, it seems to me, a further function of this type of memory is to support the representation of a self. Autobiographical memories are the things that

ground the self, and they ground it in the past. The classification of memories in this book provides a thoughtful insight into how this grounding might take place. □

Martin Conway is in the Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, UK.

Seminal work

Sperm Competition and Sexual Selection

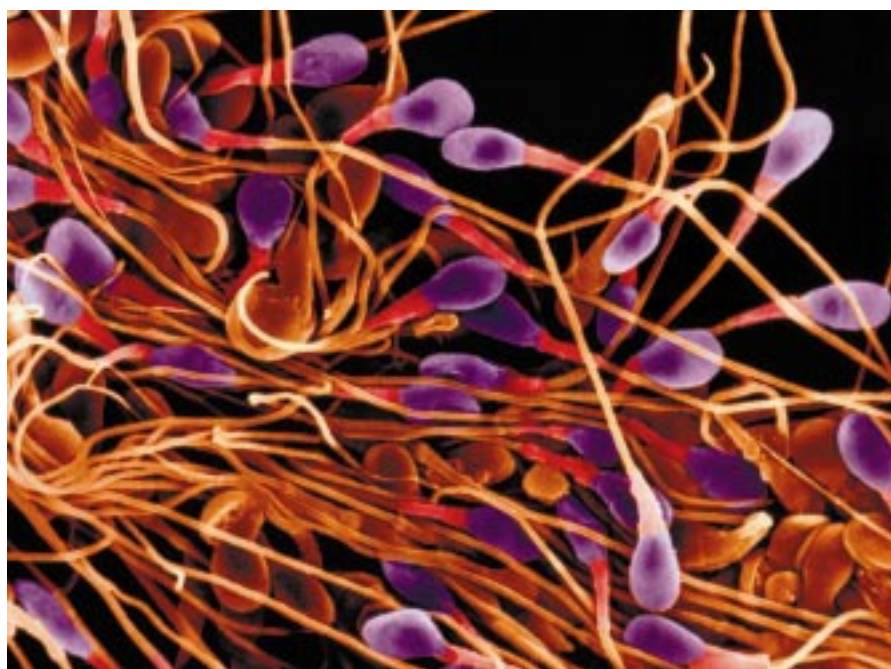
edited by T. R. Birkhead and A. P. Møller
Academic Press: 1998. 826 pp. \$59.95

Mark Ridley

Geoff Parker defines sperm competition as "competition between the sperm from two or more males for the fertilization of a given set of ova". He practically invented the subject in 1970, and research has grown explosively since then. This book describes the state of the art, succeeding the multiauthor review edited by R. L. Smith in 1984.

Since that volume was published, two big new themes have emerged. The first of these is the use of molecular markers to measure multiple paternity within a female's offspring. Paternity was almost impossible to measure in natural populations before molecular markers became available in the 1980s, but we now have bags of data, particularly for birds and mammals.

Birkhead's chapter on birds uses evidence "for over 100 species and the list continues to grow". At a crude level, this evidence is valuable because it shows the extent of 'extra-pair paternity' in apparently pair-bonded birds. Sperm competition is unimportant in genuine monogamists, and the finding (or con-



Winner took all: a crowd of human sperm no longer jockeying for position.

firmed suspicion) that up to 60% of broods in apparently monogamous species have more than one genetic father has hugely expanded the range of species in which sperm competition must operate. The same method also features in tests of subtler, deeper hypotheses.

Second, the development of statistically sound methods for studying cross-species trends has caused comparative work to take off. R. V. Short noticed in 1979 that testis size is larger in chimpanzees, in which females copulate with several males, than in other (one-male-per-female) great apes. Sperm competition will be stronger in the 'multi-male' mating system of the chimpanzee, and they have duly evolved larger sperm facto-

ries, and pump more of the stuff out.

The same relationship between testis size in a species and the number of males that a female copulates with has been rigorously demonstrated in butterflies, fish, frogs (Halliday describes how foam-nesting frogs have larger testes than other kinds), reptiles (probably), birds, marsupials and eutherian mammals, making it one of biology's most general laws. In primates, the interaction with other factors, such as the length of the breeding season, has been traced, and the components broken down, to show that males in species with multi-male mating systems produce larger numbers of sperm, store more of it, and in some cases have larger sperm.

Molecular markers have not been used to measure sperm competition in humans, but reverse comparative engineering has. Reverse comparative engineering, however, shows that human testes are on the small side, compared with our relatives, and suggests a monandrous evolutionary history as either monogamists or polygamists with one male per harem.

Research into the mechanisms of sperm competition has given us several pleasing stories to add to the whirring parts and scrapers of the damselfly penis, exposed by Waage in 1979. A male bushcricket, "using a modified subgenital plate", turns the female's reproductive tract inside-out, and then "the female consumes any stored sperm. In this way 85% of stored sperm are removed before the male transfers his spermatophore." A ghost crab "uses a glue-like substance in the seminal fluid to seal in any previously stored sperm and prevent it from fertilising". Simmons and Siva-Jothy con-

New editions

The RNA World, 2nd edn

edited by Raymond F. Gesteland, Thomas R. Cech and John. F. Atkins
Cold Spring Harbor Laboratory Press, \$129

Massive Neutrinos in Physics and Astrophysics, 2nd edn

Rabindra N. Mohapatra and Palash B. Pal
World Scientific Publishing, \$48, £33

Foundations of Earth Science, 2nd edn

Frederick K. Lutgens and Edward J. Tarbuck
Prentice Hall, £24.95, \$58

Meteorites and Their Parent Planets, 2nd edn

Harry Y. McSween, Jr.
Cambridge University Press, £40 (hbk), £14.95, \$24.95 (pbk)

Invertebrate Palaeontology and Evolution, 4th edn

E. N. K. Clarkson.
Blackwell Science, £24.95

The New Solar System, 4th edn

edited by J. Kelly Beatty, Carolyn Collins Petersen and Andrew Chaikin
Sky Publishing and Cambridge University Press, £40, \$59.95

The Insects: Structure and Function, 4th edn

R. F. Chapman
Cambridge University Press, £35, \$54.95

Universe, 5th edn

by William J. Kaufmann III and Roger A. Freedman
Freeman, \$64.95, £26.95

clude for insects that “we are still largely ignorant of the mechanisms by which non-random paternity is generated”, which may be true (it almost has to be for insects) but is too modest.

The theory of sperm competition can also be used to make sense of adaptations in individual species. This work ranges from the rigorous to the anecdotal. Parker's update of his model of optimal copula duration in the dungfly — successfully predicting the dependence of copula duration on partner body size — is at the extreme of impressive rigour. The anecdotes, however, can be equally amusing, and at least point to research problems for the future. I liked this juxtaposition: hermaphrodite slugs “frequently lose their penis during copulation. It becomes stuck in the female tract and is bitten off either by the receiver or its possessor... The presence of multiple copulation organs in some free-living flatworms has been interpreted as a way to replace copulatory complexes lost during copulation.” And then, despite 250 years of research (reviewed by Baur), we still do not know the function of the “love darts” that certain hermaphroditic snails and slugs may fire into their partners during mating.

A final theme is “cryptic female choice”, the idea that females choose among the sperm of different males after insemination. It has captured the imagination of most of the authors, and was championed in a 1996 book by Eberhard (who has a chapter on it here). The consensus is that it is one for the future. Evidence for pollen usage, in the chapter on plants by Delph and Havens, suggests one model system. Or the wonderful experiment on fruitflies by Rice may point the way.

The book is bang up to date, referring to work done as recently as 1997. Topicality may have sometimes been gained at the cost of integration. More than one chapter may allude to the same published work, but none of them explains it fully. Chapters sometimes refer to papers discussed by their authors elsewhere in the book. I'd also question the taxonomic organization, with most chapters being about a particular group. This is not how I subdivide the subject, and it is awkward that one of the most interesting recent studies on a sea squirt receives short shrift, perhaps because that obscure taxon did not deserve a whole chapter. But these criticisms are piffling and my main reaction is to praise the immense achievement.

Two of the chapters — by Gomendio, Harcourt and Roldán on mammals, and by Simmons and Siva-Jothy on insects — are expert minibooks in themselves. The book contains large appendices and tables of data; background reproductive physiology made intelligible for behavioural ecologists; intelligent suggestions for future research; and original analyses which have not been pub-

lished elsewhere. It is a one-stop first-step guide for anyone wishing to understand, add to, subtract from (critically), or multiply up (synthetically) research on sperm competition, and should be compulsory equipment in any lab that works on sexual selection. □

Mark Ridley is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Illuminating the past

Lengthening the Day: A History of Lighting Technology

by Brian Bowers

Oxford University Press: 1998. 220 pp.

\$45, £24.99

Peter T. Fox

Artificial light, like food and shelter, is a necessity for civilized life. Since Prometheus stole fire from the gods, mankind has manipulated light to ward off the dangers of the dark and to extend the hours in which vision-based activities can be pursued. Lighting adequate for such basic needs is readily provided by rather rudimentary technologies: rushes, candles, oil lamps and the like. Making utilitarian light safe, conve-

nient, cool, clean and olfactorily inoffensive at a cost affordable by the masses, however, is a much greater challenge, and has motivated many developments in lighting technology.

Yet light is much more than merely utilitarian. From the Roman era onward, luxury (*lux*) has been virtually synonymous with light (*lux*). The well-lit environment is a hallmark of personal and societal wealth. Well-balanced lighting enhances our enjoyment of visual arts, theatre, food and the face across the dinner table. Extravagant lighting displays have been crowd-pleasers for as long as there have been crowds. Competition between cities for the most artistic and impressive lighting of public buildings, bridges and roads has been in full sway for more than a century and shows no sign of abating. Thus, luxury has been an equally powerful force behind the lighting industry.

In *Lengthening the Day*, Brian Bowers gives a multi-faceted history of the development of lighting technology. The problems that had to be solved, the personalities of the inventors, the patents awarded, the impact of patents on subsequent developments, and the incremental improvements in the quality of life are thoroughly and entertainingly documented.

All manner of lights — from the gas lime-lights in theatres of the 1860s, to the first filament bulbs of Swan and Edison, to the latest



Shop around the illuminated clock: from the Roman era onwards, light and luxury have gone hand-in-hand. A well-lit environment still suggests wealth and comfort: an attempt to get us to relax and spend.

generation of induction lamps and LEDs — are explained, illustrated and critiqued. More than 130 illustrations (including seven colour plates) depict an enormous variety of lamps and ancillary technologies, including wick trimmers, filament pullers, bulb-evacuating pumps and primitive electrical generators. Nearly all the plates are either reproductions of period illustrations or photographs of museum pieces in action.

Equally lavish are the text excerpts from an impressive array of sources. Bowers and his associates must have scoured thousands of potential sources to select the scores of choice quotations with which he depicts not merely the technology of lighting, but its sociology, economics and politics.

For a book about technology, nostalgia is surprisingly pervasive. Vignettes recounting the idiosyncrasies, romance, ingenuity and inconveniences of bygone lamps provide a nicely balanced counterpoint to the technical descriptions.

Despite my unfamiliarity with non-electric lighting, I found myself gently coerced into recalling the gas-mantle lanterns of childhood camping expeditions,

Bunsen burners in high school chemistry laboratories, the gas wall fixtures long since fitted for electric bulbs in a turn-of-the-century home I once owned, and floating-wick

oil lamps in trendy restaurants.

Clearly infatuated with antique lighting systems, Bowers manages to engage the reader in his avocation by speeding through the chronology of lighting, and by moving fluidly among the many facets of each era's technology. A measure of his success at striking this delicate balance is that I finished the book wanting more, with an appetite whetted for deeper explanations of the physical principles of lighting and curious to learn the very latest developments and applications, particularly regarding the technical workhorses of our era: aeronautics, computers, entertainment and medicine. □

Peter T. Fox is in the Research Imaging Center, University of Texas Health Science Center, San Antonio, Texas 78284-6240, USA.

A thing to a void

Nothingness: The Science of Empty Space

by Henning Genz

Perseus Books: 1998. 340 pp. \$30, £20.95

Michael Redhead

If you tried to empty a box, by taking out all the material objects inside it, sucking out the air with a super-efficient pump, and cooling it towards absolute zero to eliminate thermal radiation, what would you be left with? An empty space, but is that nothing or something? If something, the box is not really empty, but if nothing, how could it be said to exist at all? These semantic puzzles about the void or vacuum intrigued the ancient Greeks. Some, like Aristotle, denied that the vacuum was a coherent concept. Others, like the ancient atomists, thought it necessary to allow the atoms to move about and change their configurations.

In this new translation of a German text published in 1994, a theoretical physicist, Henning Genz, has had the happy idea of linking these ancient discussions to the subsequent history of scientific investigation of the vacuum, such as the seventeenth-century studies of Evangelista Torricelli, Otto von Guericke and Robert Boyle, and the nineteenth-century ether theories of electrical, magnetic and optical phenomena. This leads on to modern theories of the vacuum described by quantum field theory, with exotic phenomena such as vacuum fluctuations, virtual particles, the Casimir effect, the mysterious Higgs field — supposedly responsible for spontaneous symmetry breaking of the vacuum state — and much else besides. So the vacuum of modern theoretical physics is seething with activity, and is very far removed from nothing!

While the idea of such a book was a happy one, its execution leaves much to be desired. Genz fails to cover both key issues from history and important recent developments. The history is sloppy, largely reliant on secondary sources. There is no sense of the subtlety of the arguments of the ancient Greeks. Indeed, they are made to look just stupid and

ignorant. Nearer our own times, names are misspelt and dates inaccurate — the most glaring howler is the remark that Einstein's paper on the photoelectric effect dates from 1900 (it was 1905). The translation is clumsy and in the bibliography the works of Plato and Galileo, for example, are still cited in German editions.

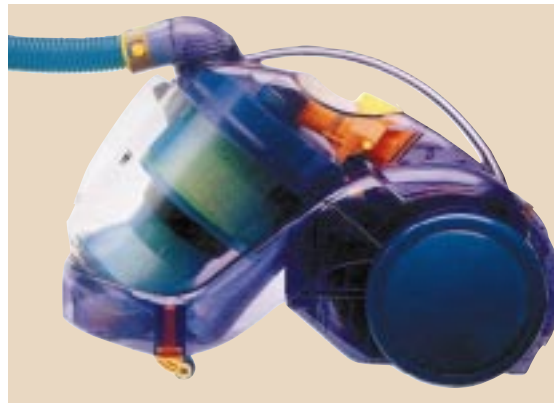
In explaining the physics, Genz uses no equations (other than $E = mc^2$) and the level of exposition is much below the standard of *Scientific American*. Some of the cosmological discussion is well done but some of the explanations are quite baffling. I defy anyone to follow the discussion of the Michelson–Morley experiment, where the motion of the source and the motion of the observer are systematically confused. There is some discussion of Mach's principle, but I was unable to discern whether the author thought that general relativity exemplified it or not. There is passing reference to vacuum solutions of the Einstein field equations, but no adequate discussion of their significance.

In explaining vacuum fluctuations, Genz deploys familiar arguments in terms of the creation and annihilation of virtual particles. But the fact that free fields also exhibit vacuum fluctuations shows that interactions are not really the source of the problem, which lies, more accurately, in the fact that the global vacuum is irreconcilable with a local vacuum in relativistic quantum field theory, the direct opposite of the situation in nonrelativistic quantum field theory.

Moreover, Genz gives no discussion of the recent literature on the Bell inequalities and the Einstein–Podolsky–Rosen argument for the incompleteness of quantum mechanics, in the context of relativistic quantum field theory. These exciting developments throw much light on the possible interpretations of vacuum fluctuations in relativistic quantum field theory.

In summary, Genz's project is thoroughly worthwhile, but it is to be hoped that someone better qualified will make a better effort on some future occasion. □

Michael Redhead is at the Centre for the Philosophy of the Natural and Social Sciences, London School of Economics, Houghton Street, London WC2A 2AE, UK.



Nature abhors a vacuum

But the British public likes this one. The popular Dyson Dual Cyclone vacuum cleaner is featured in the slightly ad-hoc selection of "clever things ... which have changed our lives over the last 100 years" in *Century Makers* by David Hillman and David Gibbs (Weidenfeld & Nicolson, £16.99, \$22.95).

A mammalian protein with specific demethylase activity for mCpG DNA

Sanjoy K. Bhattacharya*, Shyam Ramchandani*, Nadia Cervoni & Moshe Szyf

Department of Pharmacology and Therapeutics, McGill University, 3655 Drummond Street, Montreal, QC, H3G 1Y6, Canada

* These authors contributed equally to this work.

DNA-methylation patterns are important for regulating genome functions, and are determined by the enzymatic processes of methylation and demethylation. The demethylating enzyme has now been identified: a mammalian complementary DNA encodes a methyl-CpG-binding domain, bears a demethylase activity that transforms methylated cytosine bases to cytosine, and demethylates a plasmid when the cDNA is translated or transiently transfected into human embryonal kidney cells *in vitro*. The discovery of this DNA demethylase should provide a basis for the molecular and developmental analysis of the role of DNA methylation and demethylation.

Many lines of evidence indicate that DNA methylation is important in differential control of gene expression¹. The molecular mechanisms through which methylation of cytosines in the dinucleotide CpG repress transcription are now being determined²⁻⁴. The process that determines which CpGs are methylated is critical for proper development⁵ and possibly maintenance of somatic biological functions. Two enzymatic processes are involved in laying down the pattern of methylation during development, namely methylation and demethylation⁶⁻⁹. The cDNA encoding the enzyme responsible for methylation, DNA methyltransferase, has been purified and cloned from several organisms^{10,11}, but the nature of the enzymatic process responsible for demethylation has remained controversial. As the removal of methyl groups from cytosines has been presumed to be energetically unlikely, different alternative biochemical pathways have been proposed to explain active demethylation, such as pathways involving glycosylase¹²⁻¹⁴ or a nucleotide-replacement reaction^{15,28}. To resolve this issue, we need to know the molecular identity of the enzyme responsible for demethylation. We now show that mammalian cells contain an enzyme that demethylates DNA by catalysing the removal of the methyl group on the 5' position of cytosines residing in the dinucleotide sequence mCpG, releasing the methyl group in the form of methanol (our unpublished observations). However, the identification of the gene encoding demethylase and the characterization of its molecular properties will be essential for establishing its presence in vertebrate biology and for further understanding its role in development and in the formation of the DNA-methylation pattern, as well as its possible role in pathological states.

Identification of a DNA-demethylase candidate

As the purification of demethylase indicates that it may be of very low abundance (our unpublished observations), we opted to clone the cDNA encoding the demethylase by making use of the fact that the enzyme specifically demethylates methylated CG dinucleotides (our unpublished observations). We assumed that it should be able to recognize methylated CG dinucleotides. Proteins interacting with methylated DNA share a common domain, the methyl-CpG-binding domain (MBD)¹⁶. We performed a TBLASTN¹⁷ search of the dbEST database using the MBD of the methyl-CpG-binding protein MeCP2, and identified a new expressed sequence tag (EST) cDNA (from a T-cell-lymphoma *Homo sapiens* cDNA 5' end) (gb/AA361957/AA361957 EST71295) and a mouse homologue (gb/W97165/W97165 mf90g05.r, from Soares mouse embryo NbME13.5). These cDNAs encode proteins of unknown function that show homology to the MBD of MeCP2 (Fig. 1a). The same cDNA has been cloned and identified as encoding the methylated-

DNA-binding protein MBD2b (ref. 18). Alignment of our new EST with the cDNAs encoding MeCP2 and an MeCP1-associated protein revealed no homology beyond the previously characterized MBD, which is consistent with a different function for this methylated-DNA-binding protein. We cloned a 1.36-kilobase (kb) cDNA from HeLa cells. A 'virtual translation' of the cDNA identified an open reading frame (ORF) of 262 amino acids (Fig. 1b). The ORF may extend further in the 5' direction, as no in-frame stop codon was found upstream of the ATG used. A longer cDNA encoding mouse MBD2 has been cloned¹⁸.

A BLAST search^{19,20} of the candidate protein using the Predict protein server²¹ against a database of protein-domain families identified only the MBD. No other functional motifs were identified by Prosite analysis^{21,22}. This is consistent with a new function for this protein. A coiled-coil prediction²³ of the sequence identified a coiled-coil domain that is involved in protein-protein interactions²⁴.

The identified cDNA encodes a messenger RNA that is broadly expressed in human cells; this mRNA was revealed by northern blot analysis of human poly(A)⁺ mRNA (Fig. 1c) to be one major transcript of ~1.6 kb, which is close to the size of the cloned cDNA, confirming that the cloned cDNA does not represent a highly repetitive RNA, but rather a mRNA encoded by a single or low-copy-number gene.

Demethylation activity of the translated candidate cDNA

A conclusive proof for the existence of a single protein that demethylates DNA is the demonstration that an *in vitro*-translated candidate cDNA can transform a methyl cytosine to cytosine in an isolated system. We subcloned the candidate demethylase cDNA into a pcDNA3.1/His Xpress expression vector in the putative translation frame (producing pcDNA3.1His A) and in a single-base frameshift (producing pcDNA3.1His B), and the cDNA was transcribed *in vitro* and translated in the presence of [³⁵S]methionine. The resulting translation products were resolved by SDS-polyacrylamide gel electrophoresis (PAGE). Autoradiography revealed a protein of relative molecular mass 40,000 (*M_r* 40K) (Fig. 2a).

The *in vitro*-translated protein (purified on a Ni²⁺-charged agarose resin) transformed CH₃-cytosine residing in [α-³²P]deoxy (d)GTP-labelled plasmid DNA or in (methyl-dC³²pG)_n double-stranded oligomer DNA to cytosine, whereas a frameshift *in vitro*-translated demethylase did not demethylate DNA (Fig. 2b). Thus the demethylase activity is dependent on the demethylase translation product and not on a contaminating activity found in the *in vitro*-translation kit that co-purifies with the putative demethylase. These results show that the candidate cDNA encodes a *bona fide* demethylase.

Transfected demethylase cDNA demethylates DNA

The pcDNA3.1His A and pcDNA3.1His B vectors were transiently transfected into human embryonal kidney cells to test whether the cDNA would direct expression of demethylase activity in human cells. The histidine-tagged proteins were bound to Ni²⁺-agarose resin and eluted from the resin with 700–1,000 mM imidazole. The expression of the transfected demethylase was verified by western blot analysis (Fig. 3a). The transiently expressed demethylase encoded in pcDNA3.1His A transformed methylated cytosine in two different DNA substrates to cytosine (Fig. 3b, c), in a protein-dependent manner (Fig. 3b, d). The transiently transfected demethylase demethylated a methylated SK plasmid, as indicated by the sensitivity of the plasmid to the restriction enzyme *HpaII* following its incubation with demethylase but not after incubation in the presence of the protease-K-pretreated demethylase or buffer alone (Fig. 3d). The reaction showed substrate dependence and saturability (Fig. 3e). We loaded transiently expressed demethylase onto a non-denaturing glycerol gradient to determine its native relative molecular mass. Like demethylase purified from human cells (our unpublished observations), our cloned and purified

demethylase activity fractionated in the 160K–190K range (data not shown). This is consistent with self-association of our demethylase, possibly mediated by the coiled-coil domain.

A demethylase activity from human cancer cells

To determine whether a demethylase activity like the one shown by the synthetic protein could be isolated from human cells, we subjected nuclear extracts from non-small-cell human lung carcinoma A549 cells to sequential chromatography (see Methods). An active demethylase fraction transformed methyl-cytosine in a (methyl-dC³²pdG)_n double-stranded oligomer DNA to cytosine (Fig. 4a). We then subjected demethylase-treated DNA to remethylation with the CpG methyltransferase M.SssI, which can transfer a methyl group exclusively to deoxycytosine²⁵. The demethylated product of demethylase must be deoxycytosine, as it is completely remethylated by M.SssI (Fig. 4a). The identity of the demethylated product was confirmed by two-dimensional thin-layer chromatography (TLC), which showed that the product of demethylase migrated with a cold dCMP standard in both dimensions (Fig. 4b).

Demethylase did not release a nucleotide, a phosphorylated base

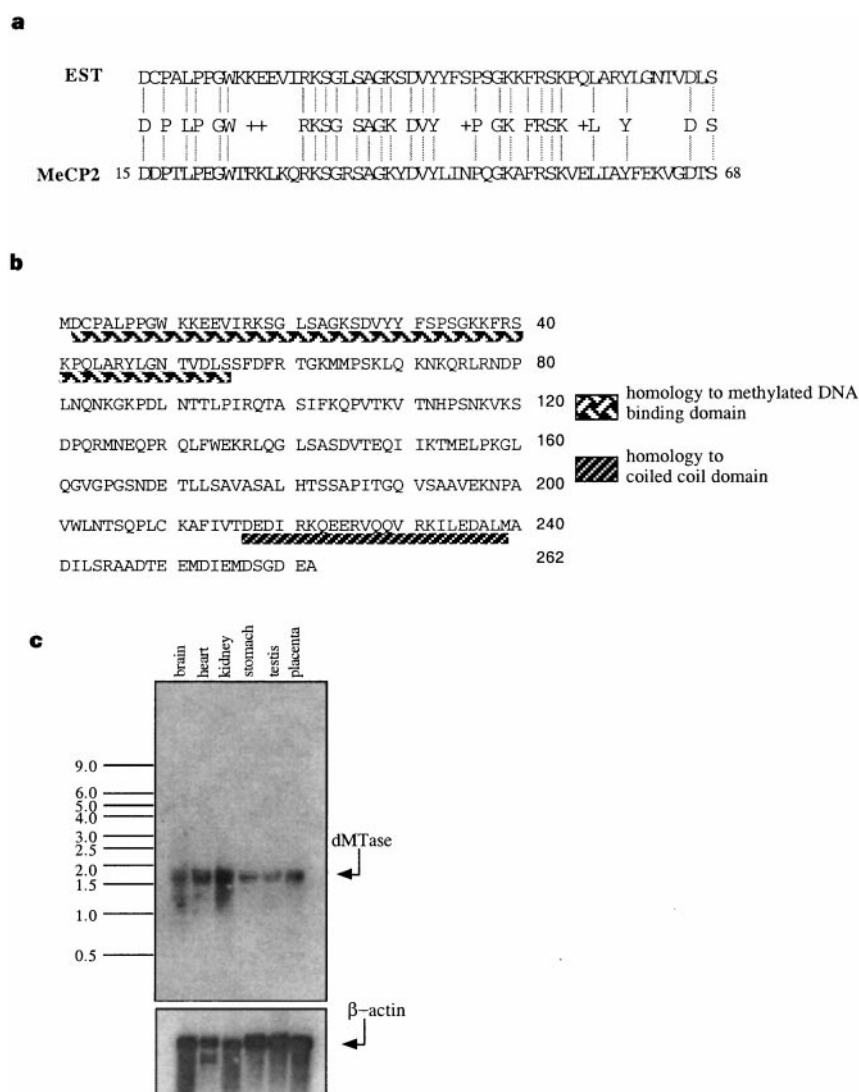


Figure 1 Human demethylase cDNA. **a**, Homology between the methylated-DNA-binding domain (MBD) of MeCP2 (amino acids 15–68) and our putative demethylase (EST). **b**, The sequence of the putative demethylase cDNA was determined and virtually translated by DNASTAR. The 262-amino-acid open reading frame is shown. The region containing homology to the MBD of MeCP2 is

highlighted at the amino-terminal of the protein sequence and a predicted coiled-coil domain is highlighted at the carboxy terminus. **c**, A northern blot analysis of 2 µg poly(A)⁺ mRNA prepared from different human tissues (Origene Technologies) hybridized with demethylase (dMTase) cDNA as a probe.

or phosphate from methylated DNA when incubated with a (^{32}P -methyl-dCpdG) $_n$ substrate that included ^{32}P located 5' to methyl-deoxycytosine, or when incubated with our standard methylated substrate in which ^{32}P was 3' to the methyl-5-deoxycytosine (Fig. 4c). Nuclear extracts that contained several glycosylases and nucleases released phosphorylated derivatives in the same assay (Fig. 4c). Demethylase transformed the methyl cytosine in the (^{32}P -methyl-dCpdG) $_n$ substrate to cytosine, as shown when the reacted DNA was digested to 5' mononucleotides (Fig. 4c, middle panel) and analysed by TLC. As this reaction does not involve release of a ^{32}P derivative (Fig. 4c, left panel), the demethylase must transform methylated cytosines in DNA to cytosines without disrupting the

integrity of the DNA substrate by glycosylase or nuclease activity. These results establish that mammalian cells express a *bona fide* demethylase activity, but it remains unknown whether the cloned cDNA encodes the activity purified from A549 cells.

Discussion

An understanding of how methylation patterns are created, maintained and modified and how they are involved in development and pathology has been stifled by the uncertainty surrounding the identity of the enzymes involved in demethylation. We have described here the cloning of a single cDNA that encodes a protein of M_r 40K in both *in vitro* translation and transient transfection

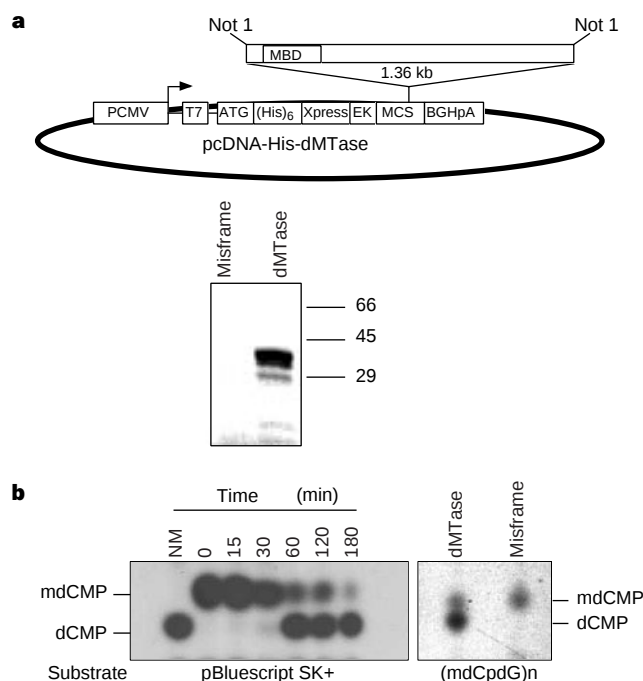


Figure 2 *In vitro* expression and demethylase activity of the human demethylase encoded by the cloned cDNA. **a**, T7-coupled histidine $_6$ tagged mammalian expression construct containing the entire cloned demethylase cDNA in-frame (dMTase) or a single-base-pair frameshift (misframe) were transcribed and translated *in vitro*, revealing an in-frame polypeptide with an apparent M_r value of ~40K (bottom). **b**, Left, demethylation of 2 ng [α - ^{32}P]dGTP-labelled methylated SK DNA by incubation with 10 μl Probond-purified (700 mM imidazole fraction) *in vitro*-translated demethylase for varying times. Plasmid pBluescript SK (5 ng) was methylated with M.SssI methylase. *In vitro*-methylated pBluescript SK was used as a template for cDNA synthesis using DNA polymerase and [α - ^{32}P]dGTP, methyl-dCTP, dTTP, dATP and hexanucleotide primers. The treated DNA was digested to 3' mononucleotides with micrococcal nuclease and subjected to TLC. The resulting methyl-dCMP (mdCMP) and dCMP (non-methylated cytosine) mononucleotides are shown. NM, non-methylated substrate control. Right, *in vitro*-translated DNA demethylase, but not misframe demethylase, demethylated cytosines in (methyl-dC 32 pdG) $_n$ ((mdCpdG) $_n$). The mononucleotide products mdCMP and dCMP are shown. kb, kilobases.

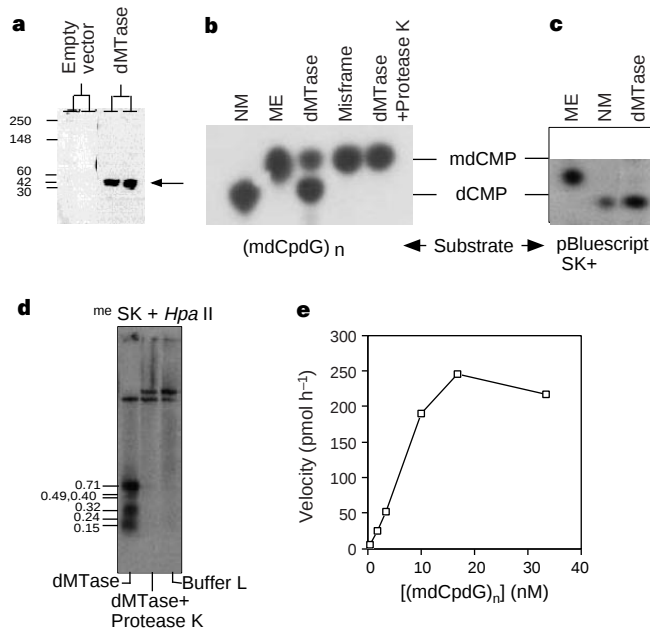


Figure 3 Expression of cloned demethylase cDNA in HEK cells. **a**, A Western blot of the purified DNA demethylase (dMTase) and expressed vector control, pcDN3.1HisC (empty vector), from Probond fraction 700 mM imidazole. The Western blot was reacted with the anti-Xpress-epitope antibody (Invitrogen). **b**, Demethylation of (methyl-dC 32 pdG) $_n$ ((mdCpdG) $_n$) DNA treated with 10 μl Probond-purified demethylase (fraction 700 mM; dMTase), Probond-purified misframe demethylase (fraction 700 mM; misframe), or Probond-purified demethylase (fraction 700 mM) pretreated with protease K (20 μg at 50°C for 2 h; dMTase + protease K). NM, non-methylated substrate control; ME, untreated methylated DNA control. The positions of methyl-dCMP (mdCMP) and dCMP (non-methylated cytosine) are indicated. **c**, Demethylation of [α - ^{32}P]dGTP-labelled SK DNA treated with 10 μl Probond-purified demethylase (fraction 700 mM) (dMTase). **d**, Southern blot analysis of *Hpa*II digestion of 5 ng methylated SK DNA (me SK) treated with 5 μl Probond-purified demethylase (fraction 700 mM; dMTase), Probond-purified demethylase (fraction 700 mM) pretreated with protease K (dMTase + Protease K), or incubation buffer alone (Buffer L). Demethylation using these concentrations (5 μl of fractionated demethylase) is not complete. Under partial demethylation conditions, some molecules are fully demethylated whereas others remain fully methylated. This is consistent with a processive mechanism for the enzyme (N.C. *et al.*, unpublished observations). **e**, Dependence of demethylation activity on substrate concentration. 1 ng methylated [dC 32 pdG] DNA was supplemented with non-radioactive, methylated [dCpdG] DNA to the indicated concentrations, incubated with 10 μl Probond-purified demethylase (fraction 700 mM) in a 100- μl reaction volume in buffer L for 3 h at 37°C, and analysed by TLC. Each pmol of DNA substrate bears 500 pmol of 5-methyl-cytosine (the average size of the double-stranded substrate is 500 base pairs). The rate of transformation of methyl-cytosine to cytosine was determined following quantification of the corresponding spots on a phosphorimager.

systems. This single polypeptide demethylates DNA in a defined *in vitro* system, providing conclusive proof for the existence of a *bona fide* demethylation machinery in mammals and presenting us with the molecular tools with which to dissect the biological functions of the demethylase activity. Although our data indicate that a single protein may bear demethylation activities, it is possible that this protein acts in a homomeric complex, as suggested by the sizing of native transiently transfected cloned cDNA.

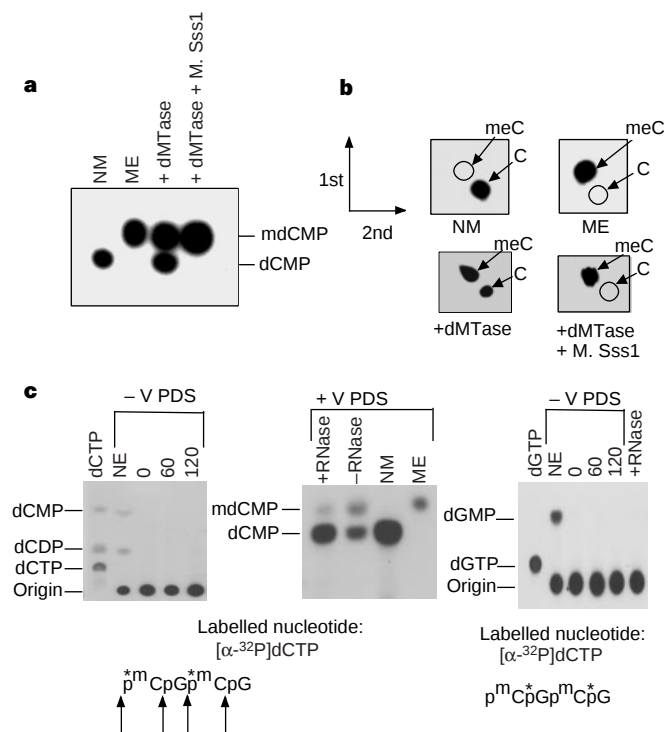


Figure 4 DNA-demethylase activity purified from human cancer cells produces cytosine and exhibits no exonuclease or glycosylase activity. **a**, DEAE-Sephacel demethylase fraction (0.5 units) was incubated with 1 ng (methyl- $C^{32}pG$)_n substrate in a 50- μ l total reaction volume (see Methods) and was divided into two 20- μ l aliquots. One aliquot was subjected to methylation with 10 units M.SssI methylase. The remethylated aliquot (+dMTase + M.SssI), the non-methylated aliquot (+dMTase), the (methyl- $C^{32}pG$)_n substrate that had not been treated with demethylase (ME) and a non-methylated ($C^{32}pG$)_n substrate (NM) were digested to produce 3' mononucleotides (methyl-dCMP (mdCMP) and dCMP) and analysed by TLC. **b**, Two-dimensional TLC analysis of the above reaction products in the presence of cold dNMP standards (solvent 1: 66 isobutyric acid:20 H₂O:1 NH₃ solvent 2: 80 (NH₄)₂SO₄:2 isopropyl alcohol:2 sodium acetate). The positions of the cold dCMP or mdCMP spots were identified by shining an ultraviolet light on the plate and are indicated in both the one-dimensional (**a**) and the two-dimensional (**b**) analysis. meC, methylated cytosine. **c**, Left, DEAE-Sephacel-purified demethylase (1 unit) was incubated with 1 ng (^{32}p -methyl-dCpG)_n substrate in 50- μ l total reaction volume containing buffer L; 3- μ l aliquots were withdrawn at different times, as indicated, and analysed by TLC for the presence of excised ^{32}P residues. As a control, the DNA was incubated for 120 min with A549 nuclear extract (NE). [α - ^{32}P]dCTP was run as a control for the absence of non-incorporated dCTP in the labelled DNA (dCTP). Middle, to determine whether demethylase demethylated the DNA in this reaction, the reaction products were digested with a 5' venom phosphodiesterase (+V PDS) after 2 h of incubation in the presence or absence of 100 μ g ml RNase as indicated. (^{32}p -methyl-dCpG)_n that was incubated in the absence of demethylase (ME) and non-methylated (^{32}p -dCpG)_n (NM) were run as controls. Right, DNA demethylase (dMTase) was incubated with (methyl-d $C^{32}pG$)_n substrate, and aliquots were withdrawn at different times and subjected to TLC analysis (shown) and venom phosphodiesterase treatment (data not shown), as described for the middle panel. At the bottom are shown labelled DNA substrates and cleavage sites; asterisks indicate the labelled phosphate groups, and arrows indicate sites of cleavage by venom phosphodiesterase.

We have shown that a general methyl-dCpG demethylase exists and that this new enzyme can perform the activities that are necessary for demethylation *in vivo*. Our results identify a new enzyme and biochemical reaction that have not yet been described in any organism. The results indicate that the methyl group may be removed in this reaction. Further experiments will be required to unravel the mechanism of the reaction and identify the leaving group. We propose that the reactants of the new reaction are H₂O and methyl-dCpG-bearing DNA and that the products of the reaction are dCpG-bearing DNA and methanol. We have shown by TLC (Figs 2b, 3b, c and 4) and restriction enzyme analysis (Fig. 3d) that cytosine is the product of the demethylase reaction. The DNA is intact following the demethylase reaction, as shown by the lack of exonuclease activity (Fig. 4c). Our unpublished preliminary data obtained using vitalization assays, gas chromatography and mass spectrometry indicate that the leaving group may be methanol. As cleavage of a carbon-carbon bond requires high energy, demethylation of 5-methyl-cytosine *per se* has been believed to be thermodynamically unfavourable. If the products of the demethylation reaction are non-methylated cytosine and methanol, the reaction is thermodynamically favourable.

Our data indicate that, like DNA methyltransferase, which exhibits a general recognition of CG dinucleotides, demethylase is an enzyme that can generally recognize methylated CGs. The presence of demethylase mRNA in somatic cells (Fig. 1c) may allow for plasticity in the methylation pattern even in fully developed cells. Altering the level of demethylase in response to extracellular signals²⁶ or modulating its accessibility to specific sites as a result of the action of *trans*-acting factors can allow for modulation of the covalent modification pattern of the genome in response to physiological signals, and might allow for reversible modulation of gene expression by methylation²⁷.

Although further experiments will be required to demonstrate the role of demethylase in development and in somatic and cancer tissues, our data provide conclusive evidence that demethylase exists and reveal its primary molecular characteristics. The cloned cDNA will enable the further dissection of the roles of demethylase. □

Methods

Purification of demethylase from cells and tissues. Nuclear extracts were prepared from A549 (ATCC) cultures at near confluence as described²⁶. A freshly prepared nuclear extract (1 ml, 6 mg) was diluted to a conductivity equivalent to 0.2 M NaCl in buffer L (10 mM Tris-HCl, pH 7.5, 10 mM MgCl₂) and applied onto a DEAE-Sephadex A-50 column (Pharmacia) (2.0 $\times$ 1 cm) that had been pre-equilibrated with buffer L at a flow rate of 1 ml min⁻¹. Following a 15-ml wash with buffer L, proteins were eluted with a 5-ml linear gradient of NaCl (0.2–5.0 M). 0.5-ml fractions were collected and assayed for demethylase activity. Demethylase eluted between 4.9 and 5.0 M NaCl. Demethylase was purified by 8,000-fold after the DEAE step. Active DEAE-Sephadex column fractions were pooled, adjusted to 0.2 M NaCl by dilution and loaded onto an SP-Sephacel column (Pharmacia) (2.0 $\times$ 1 cm). After washing the column, the proteins were eluted with 5 ml of a linear NaCl gradient (0.2–5.0 M). 0.5-ml fractions were collected and assayed for demethylase activity. Active fractions were pooled, adjusted to 0.2 M NaCl by dilution and applied onto a protein Q-Sephacel column (Pharmacia) (2.0 $\times$ 1 cm) and proteins were eluted as described. The demethylase activity eluted around 4.8–5.0 M NaCl. The pooled fractions from the protein Q-Sephacel column were loaded onto a 2.0 cm $\times$ 2.0 cm DEAE-Sephacel column (Pharmacia) and eluted with 10 ml buffer L. The activity was detected at fraction 4, which is very near the void volume. A batch of 20 purified column fractions was pooled and subjected to cold vacuum evaporation, followed by dialysis in one litre of buffer L (no salt) at 4 °C, with three changes at 6-h intervals. One unit of demethylase activity was defined as the amount of purified enzyme that produced 1 picomol of cytosine using 1 ng methylated ^{32}p CpG substrate in 1 h at 37 °C.

Assay for demethylase activity. We assayed the conversion of methyl-dCMP as described²⁶. 1 ng of [α - ^{32}P]-labelled, fully methylated poly[methyl-d $C^{32}pG$]_n or a

control, non-methylated poly[dC³²pG]_n substrate was prepared as described²⁶, reacted in a 50-μl reaction volume in buffer L containing 10 μg RNase (all demethylase reactions were done in the presence of RNase unless indicated otherwise) with the relevant demethylase fraction for 3 h at 37°C, purified by phenol/chloroform extraction, and digested by micrococcal nuclease (Pharmacia) (100 μg at 10 μl) and incubated with calf spleen phosphodiesterase (Boehringer) (2 μg; to produce 3' mononucleotides) for 15 h at 37°C. The digestion products were loaded onto a TLC plate (Kodak, 13255 Cellulose), developed in medium containing 132 ml isobutyric acid:40 ml water:4 ml ammonia solution, and autoradiographed.

Coupled *in vitro* transcription and translation. Demethylase cDNA was cloned from a HeLa cell cDNA library in λTriplEx phage (Clontech) according to standard procedures. The positive insert in the pTriplEx plasmid was excised from the phage according to manufacturer's protocols and the identity of the insert was verified by DNA sequencing. The insert was excised by *NotI* restriction and subcloned into pcDNA3.1/His XpressA vector (Invitrogen) to generate histidine-tagged demethylase and pcDNA3.1/His XpressB to generate a single-base frameshift (misframe).

The plasmids encoded by the pcDNA 3.1/His Xpress demethylase constructs were transcribed and translated by coupled transcription–translation using Promega T7 coupled-TNT rabbit reticulocyte lysate kit (according to the manufacturer's protocol). The translation products were bound to a Probond nickel column (Invitrogen) and demethylase was eluted, according to the manufacturer's protocol, with increasing concentrations of imidazole in the range 50–1,000 mM (100 μl per step). The imidazole-eluted fractions were dialysed against buffer L as described. To assay demethylase activity, four *in vitro*-translation reactions were pooled and purified as above.

Transfection and expression of demethylase in vertebrate cells. 10 μg of either pcDNA3.1/His XpressA (demethylase) or pcDNA3.1/His Xpress B (misframe demethylase) were introduced into HEK 293 cells (ten 100-mm-diameter plates per plasmid at early log phase density) using the calcium phosphate precipitation method. Total cell extracts were prepared 48 h after transfection, bound to a Probond nickel resin (Invitrogen) and eluted with 500 μl of increasing concentrations of imidazole (50–1,000 mM) as recommended by the manufacturer. The imidazole-eluted fractions were dialysed against buffer L.

Received 12 October 1998; accepted 11 January 1999.

1. Siegfried, Z. & Cedar, H. DNA methylation: a molecular lock. *Curr. Biol.* **7**, 305–307 (1997).
2. Nan, X., Campoy, F. J. & Bird, A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* **88**, 471–481 (1997).
3. Nan, X. S. *et al.* Transcriptional repression by the methyl-CpG-binding protein MECP2 involves a histone deacetylase complex. *Nature* **393**, 386–389 (1998).

4. Jones, P. L. *et al.* Methylated DNA and MECP2 recruit histone deacetylase to repress transcription. *Nature Genet.* **19**, 187–191 (1998).
5. Li, E., Bestor, T. H. & Jaenish, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
6. Razin, A. & Riggs, A. D. DNA methylation and gene function. *Science* **210**, 604–610 (1980).
7. Brandeis, M., Ariel, M. & Cedar, H. Dynamics of DNA methylation during development. *Bioassays* **15**, 709–713 (1993).
8. Kafri, T., Gao, X. & Razin, A. Methanistic aspects of genome-wide demethylation in the preimplantation mouse embryo. *Proc. Natl Acad. Sci. USA* **90**, 10558–10562 (1993).
9. Monk, M., Boubelik, M. & Lehnert, S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**, 371–382 (1987).
10. Bestor, T., Laudano, A., Mattaliano, R. & Ingram, V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J. Mol. Biol.* **203**, 971–983 (1988).
11. Bestor, T. H. & Verdine, G. L. DNA methyltransferases. *Curr. Opin. Cell Biol.* **6**, 380–389 (1994).
12. Razin, A. *et al.* Replacement of 5-methylcytosine by cytosine: a possible mechanism for transient DNA demethylation during differentiation. *Proc. Natl Acad. Sci. USA* **83**, 2827–2831 (1986).
13. Vairapandi, M. & Duker, N. J. Enzymic removal of 5-methylcytosine from DNA by a human glycosylase. *Nucleic Acids Res.* **21**, 5323–5327 (1993).
14. Jost, J. P., Siegmund, M., Sun, L. J. & Leung, R. Mechanisms of demethylation in chicken embryos—purification and properties of a 5-methylcytosine-DNA glycosylase. *J. Biol. Chem.* **270**, 9734–9739 (1995).
15. Weiss, A., Keshet, I., Razin, A. & Cedar, H. DNA demethylation *in vitro*: involvement of RNA. *Cell* **86**, 709–718 (1996).
16. Cross, S. H., Meehan, R. R., Nan, X. & Bird, A. A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nature Genet.* **16**, 256–259 (1997).
17. Boguski, M. S., Lowe, T. M. & Tolstoshev, C. M. dbEST-database for expressed sequence tags. *Nature Genet.* **4**, 332–373 (1995).
18. Henrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **18**, 6538–6547 (1998).
19. Corpet, F., Gouzy, J. & Kahn, D. The ProDom database of protein domain families. *Nucleic Acids Res.* **26**, 323–326 (1998).
20. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
21. Bairoch, A., Bucher, P. & Hofman, K. The PROSITE database, its status in 1997. *Nucleic Acids Res.* **25**, 217–221 (1997).
22. Rost, B. PHD: predicting one-dimensional protein structure by profile based neural networks. *Methods Enzymol.* **266**, 525–539 (1996).
23. Lupas, A., Van Dyke, M. & Stock, J. Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164 (1991).
24. O'Shea, E. K., Rutkowski, R. & Kim, P. S. Evidence that the leucine zipper is a coiled coil. *Science* **243**, 538–542 (1989).
25. Nur, I. *et al.* Prokaryotic and eukaryotic traits of DNA methylation in spiroplasmas (mycoplasmas). *J. Bacteriol.* **164**, 19–24 (1985).
26. Szyf, M., Theberge, J. & Bozovic, V. Ras induces a general DNA demethylation activity in mouse embryonal P19 cells. *J. Biol. Chem.* **270**, 12690–12696 (1995).
27. Szyf, M. DNA methylation properties: consequences for pharmacology. *Trends Pharmacol. Sci.* **7**, 233–238 (1994).
28. Swisher, J. A., Rand, E., Cedar, H. & Pyle, A. M. Analysis of putative RNase sensitivity and protease insensitivity of demethylation activity in extracts from rat myoblasts. *Nucleic Acids Res.* **26**, 5573–5580 (1998).

Acknowledgements. We thank B. ten Oever and H.-Y. Shiu for technical assistance and O. A. Mamer for help with mass spectrometry. This research was supported by the National Cancer Institute of Canada.

Correspondence and requests for materials should be addressed to M.S. (e-mail: mszyf@pharma.mcgill.ca). The DNA demethylase cDNA sequence has been deposited with GenBank under accession number AF 072242.

Voluminous volcanism on early Mars revealed in Valles Marineris

Alfred S. McEwen*, Michael C. Malin†, Michael H. Carr‡ & William K. Hartmann§

* Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

† Malin Space Science Systems, San Diego, California 92191, USA

‡ US Geological Survey, Menlo Park, California 94025, USA

§ Planetary Science Institute, Tucson, Arizona 85705, USA

The relative rates and importance of impact cratering, volcanism, erosion, and the deposition of sediments to the early geological history of Mars are poorly known. That history is recorded in the upper crust of the planet, which is best exposed along the 4,000-km-long canyon system called Valles Marineris. Previous studies of the stratigraphy of this region have assumed that it consists of megabreccia and fractured bedrock resulting from impacts, overlain by or interbedded with relatively thin layers of lava, and with the layering restricted to the uppermost level of the crust^{1–6}. Here we report new high-resolution images that reveal ubiquitous horizontal layering to depths of at least 8 km in the canyons.

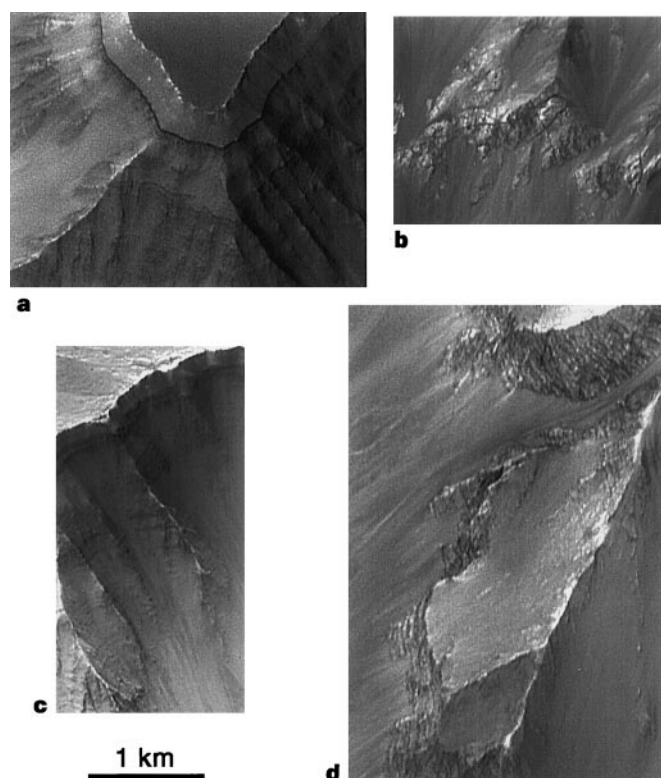


Figure 1 Portions of MOC images of wall rock in Valles Marineris. **a, c**, Exposures from near the top of the canyon; **b**, the only example of bedrock in which no layering is seen, near the bottom of the canyon; and **d**, a layered exposure from near the bottom of the canyon. Panels **a** and **c** show the uppermost cap rock, ~400 m thick throughout the length of Valles Marineris, which forms a prominent cliff that is stronger and more resistant to erosion than the underlying layers. The difference in mechanical properties could be due to alteration or cementation of upper or lower layers^{5,26–27} or the cap rock may be an especially thick flow unit. The layered block at the top of **d** appears to be tilted. Panels **a** and **c** were extracted from MOC image no. 8003; **d** is from image no. 8403 (see Table 1).

Megabreccia should be only coarsely layered and fractured bedrock should be unlayered, so these observations indicate that volcanic or sedimentary processes were much more important in early martian history than previously believed. Morphological and compositional data suggest that the layers were formed mainly by volcanic flood lavas. Mars was therefore probably very volcanically active during at least the first billion years and after the period when the heaviest impact bombardment had ended.

The Mars Orbiter Camera (MOC: refs 7, 8) acquired north-south-trending image strips over portions of the Valles Marineris canyon system during 16 out of the first 134 orbits of the Mars Global Surveyor (MGS) spacecraft⁹ (Table 1 and Figs 1 and 2). The spatial resolution varies from 4 to 10 m per picture element (pixel), so features larger than about 20 m can be discerned. The dominant morphology of the chasm walls consists of steep branching spurs and intervening gullies^{1,10}. At most locations imaged at better than 10 m per pixel, bright-dark bands 20–100 m wide are seen in the spurs. The bands usually appear to follow topographic contour lines and thus are most likely to consist of horizontal layers varying from ~5 to 50 m in thickness (assuming 10–30° slopes). The banding consists of alternate dark cliffs (or steep slopes) and brighter, shallower intervening slopes (Fig. 1). The brighter slopes may result from coatings of aeolian dust. Layering extends to the deepest bedrock exposures that have been imaged by the MOC, as much as 8 km below the plateau surface (Table 1). We have seen what looks like massive fractured bedrock at only one location (Fig. 1b).

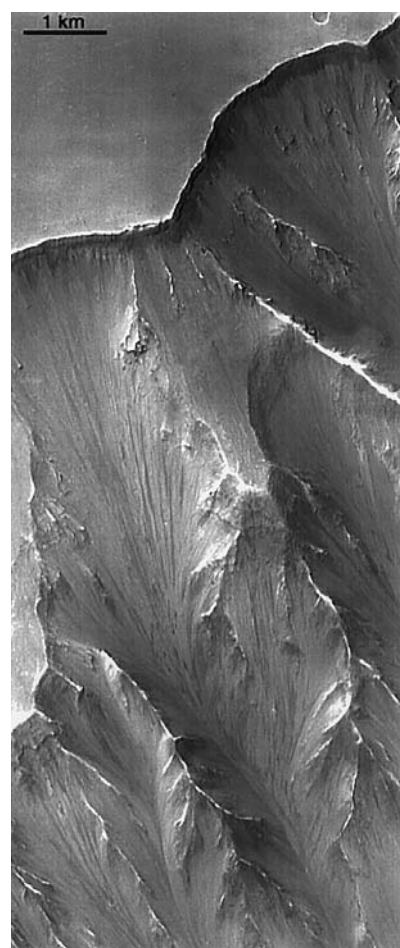


Figure 2 Portion of MOC image no. 7906. This location is in the eastern end of Valles Marineris, far from the Tharsis volcanoes near the western end, yet the canyon walls are layered down to the deepest exposures.

Table 1 MOC observations of crustal layering in Valles Marineris

Image numbers	Chasma name	Pixel scale (m)	Lat., Long. (deg)*	Thickness of layering (km)†	Geological age of top layer
1303	Ius	9.6 × 6.4	−6.6, 90.3	3–4	Upper Hesperian
4203	Ophir	6.3 × 4.2	−3.5, 72.1	5	Lower Hesperian
5103, 4, 5	Ius	6.5 × 4.3	−6, 90.8	3–4	Upper Hesperian
6303, 4, 5	East Candor	6.7 × 4.5	−6.6, 68.9	8	Lower Hesperian
6903, 4, 5	Ius	8.9 × 5.9	−7.8, 82.3	6	Upper Hesperian
7904, 5, 6	Ganges/Capri/Eos	11 × 6	−8.2, 41.9	3–4	Middle Noachian
8003	Coprates	7.2 × 4.8	−14.4, 55.8	4	Lower Hesperian
8105	Coprates	8.2 × 4.7	−11.4, 66.5	5	Lower Hesperian
8203	Ophir	6.8 × 4.6	−3.3, 72.3	3–4	Lower Hesperian
8303–4	Hebes	8.0 × 4.6	−1, 75.7	3–4	Lower Hesperian
8305	West Candor	8.0 × 4.6	−4.7, 76.0	5	Lower Hesperian
8403	West Candor	6.9 × 4.6	−5.1, 74.9	4–5	Lower Hesperian
8505	Ophir	6.9 × 4.6	−4.5, 70.1	5–6	Lower Hesperian
8506	East Candor	6.9 × 4.6	−5.6, 70.4	6	Lower Hesperian
8606	Coprates	8.4 × 4.8	−12.3, 62.6	5	Lower Hesperian
8706	Ganges	6.9 × 4.6	−6.1, 49.5	3	Lower Hesperian
8805	Hydraotes Chaos	6.9 × 4.6	−1.4, 33.0	2	Upper Noachian
9803	Ophir	9.4 × 6.3	−3.3, 72.7	4	Lower Hesperian
9804	Melas	9.9 × 6.6	−13.9, 73.8	4	Lower Hesperian

* As °S, °W.

† Thicknesses estimated from US Geological Survey topographic maps²⁴.

Most of this layering has not been seen before the MGS survey. But we note that layering has been seen in Viking and Mariner 9 images of the Valles Marineris interior deposits, formed within the canyons¹; and in the pre-existing wall rock, layering has been seen previously in the uppermost level (<1 km deep), and thick or patchy dark layers have been seen in Coprates chasm and the north wall of Ophir chasm¹¹.

The layers seen in MOC images of the Valles Marineris probably belong to the Upper Noachian stratigraphic series, thought to have ended 3.5–4.3 Gyr ago². Geological units on Mars are assigned to

one of three main time-rock units: Noachian (earliest, heavily cratered), Hesperian (intermediate) and Amazonian (youngest, lightly cratered). A period of heavy impact bombardment characterized the earliest histories of the planets, perhaps representing the final stages of planet formation. The layers in Valles Marineris underlie surface units mapped as Hesperian and Upper Noachian^{12,13} (Table 1). In one location they underlie a unit mapped as Middle Noachian (Fig. 2), but the low density of craters on the plateau suggests that the unit is in fact younger. The layers must postdate the heavy bombardment of the early-to-middle Noachian, or they would be extensively disrupted. However, there may be a selection effect in places, whereby large impacts localize regions of collapse so that the high-standing plateau is dominated by intercrater plains.

We interpret the strata as probably consisting of volcanic flows. The overlying plains probably consist of volcanic flows, on the basis of their morphology² and spectral properties¹⁴, and the layer thicknesses and cliff-forming topography are typical of continental flood basalts such as the Columbia River Basalt group¹⁵ (Fig. 3). Dark layers previously identified have spectral and other properties consistent with mafic glass¹¹, and the wall rocks in general are rich in pyroxene^{14,16}, a common igneous mineral. Voluminous volcanism is plausible given the high heat flows expected on early Mars¹⁷.

The presence of the layering under Upper Noachian terrains in the eastern Valles Marineris (see, for example, Fig. 2) suggests that the layers could also contain sediments that were deposited by water because many Noachian terrains are dissected by valley networks¹⁸. Studies of degraded craters in the ancient highlands have concluded that only ~200 m of highland material has been eroded and locally redistributed¹⁹, so perhaps there are thin layers of sedimentary deposits interbedded with the lava flows. We cannot rule out the possibility that the layers are largely sedimentary rock, but we consider the volcanic interpretation much more likely.

What happened to the megaregolith that we expected to see, especially under the Upper Noachian surfaces? Impacts must have been important, but their contribution to the upper crust depends on competing processes. It seems that the volcanic effusion rate was sufficiently high for layered lavas to dominate this crustal section, and these lavas must have been emplaced after the period of heaviest bombardment. Presumably megabreccia exists at some depth under the stack of largely undisturbed lavas.

How voluminous was this early volcanism? The areal extent of the layered deposits spans at least the entire Valles Marineris, ~4,000 km long. The surrounding plateau surface is covered by volcanic plains from Upper Noachian to Upper Hesperian age². If all martian plains younger than Middle Noachian are underlain by

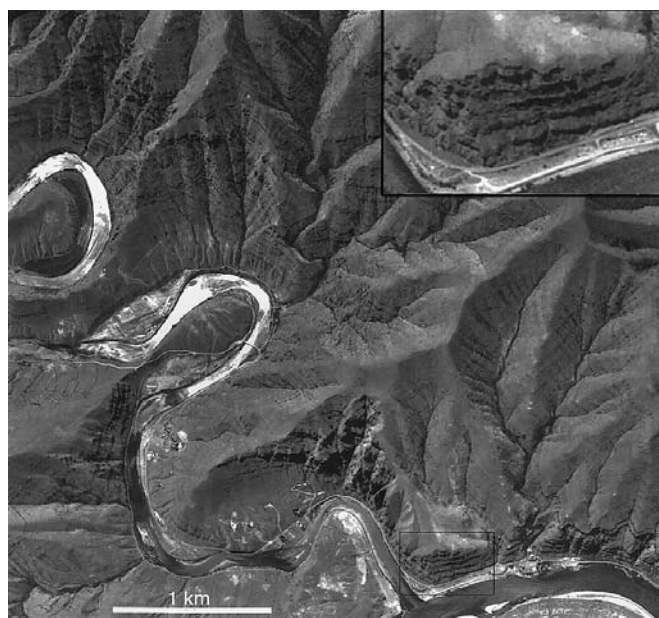


Figure 3 Aerial photograph of part of the Columbia River Basalt group in the northwestern USA; inset is an enlargement to show the morphology of the layers. The photograph was digitized at a scale of 1.66 m per pixel, similar to the resolution achievable by the MOC in a 400-km circular orbit. Continental flood basalts form layers typically 10–100 m thick, each composed of one or more flows from a single eruption. Each layer is usually separated by a thin weathered horizon. The flow layers appear to be hundreds of kilometres long. Where exposed on cliffs and hillsides, the basalts form alternate dark cliffs and brighter slopes. Although once thought to be emplaced by rapid, turbulent flow, more recent work¹⁵ favours emplacement via prolonged eruptions, with an average effusion rate of ~4,000 m³ s^{−1}.

lavas 5 km thick, then our estimates of the total quantity of lava erupted over the past ~ 4 Gyr (including volcanic edifices) must be revised from the $7 \times 10^7 \text{ km}^3$ calculated previously²⁰ to $5 \times 10^8 \text{ km}^3$. Alternatively, the lava could be less extensive if the formation of Valles Marineris was intimately associated with the presence of a thick lava sequence, which thins rapidly away from the canyons. If deep layering is confined to the area of a rectangle enclosing the canyons ($\sim 4 \times 10^6 \text{ km}^2$) and is 10 km thick, then the volume is $4 \times 10^7 \text{ km}^3$. This alone greatly exceeds a previous estimate of $8 \times 10^6 \text{ km}^3$ of magma extruded in the Late Noachian²⁰.

We conclude that volcanism on early Mars was probably much more voluminous than previously documented, and that it must have affected the climate and near-surface environment. Pollack *et al.*²¹ proposed that a warm, wet climate on early Mars was sustained by a thick CO_2 atmosphere, which must be continuously resupplied or recycled to balance loss of CO_2 to carbonates. Two mechanisms for recycling the CO_2 have been proposed: extensive volcanism²¹ and impacts²². If the recycling was mainly from impacts, then the warm, wet conditions corresponded to the time (on Earth) of heavy bombardment and the impact frustration of life²³. Extensive volcanism on Mars could have maintained a thick atmosphere for a significant period of time after the heavy bombardment²¹. The layers seen by MOC provide evidence for voluminous volcanism; but a thick atmosphere could have been sustained only if sufficient carbonates exist in the crust of Mars, which has not yet been confirmed¹⁶. □

Received 16 September; accepted 21 December 1998.

- Lucchitta, B. K. *et al.* in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. M. & Matthews, M. S.) 453–492 (Univ. Arizona Press, Tucson, 1992).
- Tanaka, K. L., Scott, D. H. & Greeley, R. in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. M. & Matthews, M. S.) 345–382 (Univ. Arizona Press, Tucson, 1992).
- Clifford, S. M. A model for the hydrologic and climatic behavior of water on Mars. *J. Geophys. Res.* **98**, 10973–11016 (1993).
- Tanaka, K. L. & Golombek, M. P. Martian tension fractures and the formation of graben and collapse features at Valles Marineris. *Proc. Lunar Planet. Sci. Conf.* **19**, 383–396 (1989).
- Davis, P. A. & Golombek, M. P. Discontinuities in the shallow Martian crust at Lunae, Syria, and Sinai Plana. *J. Geophys. Res.* **95**, 14231–14248 (1990).
- Zuber, M. T. & Aist, L. L. The shallow structure of the Martian lithosphere in the vicinity of the ridged plains. *J. Geophys. Res.* **95**, 14215–14230 (1990).
- Malin, M. C. *et al.* Mars Observer Camera. *J. Geophys. Res.* **97**, 7699–7718 (1992).
- Malin, M. C. *et al.* Early views of the Martian surface from the Mars Orbital Camera of Mars Global Surveyor. *Science* **279**, 1681–1685 (1998).
- Albee, A. L., Palluconi, F. D. & Arvidson, R. E. Mars Global Surveyor mission: Overview and status. *Science* **279**, 1671–1672 (1998).
- Lucchitta, B. K. Morphology of chasma walls, Mars. *J. Res. US Geol. Surv.* **6**, 651–662 (1978).
- Geissler, P. E., Singer, R. B. & Lucchitta, B. K. Dark materials in Valles Marineris: Indications of the style of volcanism and magmatism on Mars. *J. Geophys. Res.* **95**, 14399–14413 (1990).
- Scott, D. H. & Tanaka, K. L. *Geologic Map of the Western Equatorial Region of Mars, Scale 1:15,000,000* (Misc. Inv. Ser. Map 1-1802-A, US Geol. Surv., Denver, 1986).
- Witbeck, N. E., Tanaka, K. E. & Scott, D. H. *Geologic Map of the Valles Marineris Region of Mars, Scale 1:2,000,000* (Inv. Ser. Map 1-2100, US Geol. Surv., Denver, 1991).
- Erard, S. *et al.* Spatial variations in composition of the Valles Marineris and Isidis Planitia regions of Mars derived from ISM data. *Proc. Lunar Planet. Sci. Conf.* **21**, 437–456 (1991).
- Self, S., Thordarson, T. & Keszthelyi, L. in *Large Igneous Provinces* (eds Mahoney, J. J. & Coffin, M. E.) 381–410 (Am. Geophys. Union, Washington, D. C., 1997).
- Christensen, P. R. *et al.* Results from the Mars Global Surveyor thermal Emission Spectrometer. *Science* **279**, 1692–1698 (1998).
- Schubert, G., Solomon, S. C., Turcotte, D. L., Drake, M. J. & Sleep, N. H. in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. M. & Matthews, M. S.) 147–183 (Univ. Arizona Press, Tucson, 1992).
- Carr, M. H. *Water on Mars* (Oxford Univ. Press, New York, 1996).
- Craddock, R. A., Maxwell, T. A. & Howard, A. D. Crater morphometry and modification in the Sinus Sabaeus and Margaritifer Sinus regions of Mars. *J. Geophys. Res.* **102**, 13321–13340 (1997).
- Greeley, R. & Schneider, B. D. Magma generation on Mars: Amounts, rates, and comparisons with Earth, Moon, and Venus. *Science* **254**, 996–998 (1991).
- Pollack, J. B., Kasting, J. F., Richardson, S. M. & Poliakov, K. The case for a wet, warm climate on early Mars. *Icarus* **71**, 203–224 (1987).
- Carr, M. H. Recharge of the early atmosphere of Mars by impact-induced release of CO_2 . *Icarus* **79**, 311–327 (1989).
- Maher, K. A. & Stevenson, D. J. Impact frustration of the origin of life. *Nature* **331**, 612–614 (1988).
- Topographic Maps of the Polar, Western, and Eastern regions of Mars* (Misc. Inv. Ser. Map 1-2160, US Geol. Surv., Denver, 1991).
- Fanale, F. P. Martian volatiles: Their degassing history and geochemical fate. *Icarus* **28**, 179–202 (1976).
- Soderblom, L. A. & Wenner, D. B. Possible fossil water liquid–ice interfaces in the Martian crust. *Icarus* **34**, 622–637 (1978).
- Treiman, A. H., Fuks, K. H. & Murchie, S. Diagenetic layers in the upper walls of Valles Marineris, Mars: Evidence for drastic climate change since the mid-Hesperian. *J. Geophys. Res.* **100**, 26339–26344 (1995).

Acknowledgements. We thank L. Keszthelyi for discussions, and M. T. Zuber and N. G. Barlow for comments on the manuscript. This work was supported by the MGS project.

Correspondence and requests for materials should be addressed to A.S.M. (e-mail: mcewen@lpl.arizona.edu).

Evidence for recent volcanism on Mars from crater counts

William K. Hartmann*, Michael Malin†, Alfred McEwen‡, Michael Carr§, Larry Soderblom||, Peter Thomas¶, Ed Danielson#, Phillip James* & Joseph Veverka†

* Planetary Science Institute, Tucson, Arizona 85705, USA

† Malin Space Science Systems, San Diego, California 92191, USA

‡ Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

§ US Geological Survey, Menlo Park, California 94025, USA

|| US Geological Survey, Flagstaff, Arizona 86001, USA

¶ Cornell University, Ithaca, New York 14853, USA

California Institute of Technology, Pasadena, California 91125, USA

* University of Toledo, Toledo, Ohio 43606, USA

Impact craters help characterize the age of a planetary surface, because they accumulate with time. They also provide useful constraints on the importance of surface erosion, as such processes will preferentially remove the smaller craters. Earlier studies of martian crater populations revealed that erosion and dust deposition are important processes on Mars^{1–6}. They disagreed, however, on the age of the youngest volcanism^{7,8}. These earlier studies were limited by image resolution to craters larger than a few hundred metres in diameter. Here we report an analysis, using new images obtained by the Mars Global Surveyor spacecraft, of crater populations that extend the size distribution down to about 16 m. Our results indicate a wide range of surface ages, with one region—lava flows within the Arsia Mons caldera—that we estimate to be no older than 40–100 million years. We suggest that volcanism is a continuing process on Mars.

The distribution of crater numbers versus crater diameter on lunar lava plains, called the ‘production function’, represents the shape of the population of craters being produced on the moon in current geological time. Its shape is well determined^{9–11}. Our initial step was to test whether the production function observed on young, well-preserved surfaces on Mars is the same as that found on the Moon. This result has been found for craters larger than 1 km in diameter, but has not been well tested for the steep branch below 1 km (ref. 7) (see figures). Dashed reference lines in the shape of the lunar production function are shown in each of the crater count diagrams in this Letter, along with an upper solid line that marks the crater density on the most heavily cratered surfaces in the Solar System, dated about 4.0 Gyr old on the Moon, and believed to mark the saturation equilibrium condition where new craters erase old craters^{10,12}. These reference lines allow the comparison of the Mars craters with the lunar production function.

Here we report our analyses of images obtained by the Mars Orbiter Camera (MOC) on board the Mars Global Surveyor spacecraft. Martian crater counts obtained from the MOC images are a significant advance over previously published data, and the images reveal the importance of mobile dust in shaping the martian landscape and softening the profile of craters^{13,14}. In general, our procedure is to count all craters but avoid areas with obvious clusters of small secondary ejecta craters. To study the crater size distribution in a relatively young area, we chose a MOC image that crosses a strip of the floor of the summit caldera of the very young Tharsis volcano, Arsia Mons. Figure 1a shows some of this surface, and the crater counts obtained on Arsia Mons and its summit caldera are given in Fig. 1b. The largest crater within the summit caldera is barely 1 km across, and so the counts are extended to larger sizes (using open symbols) with additional counts from the flanks of Arsia Mons. These counts (caldera and whole volcano) each appear consistent with the shape of the lunar crater diameter

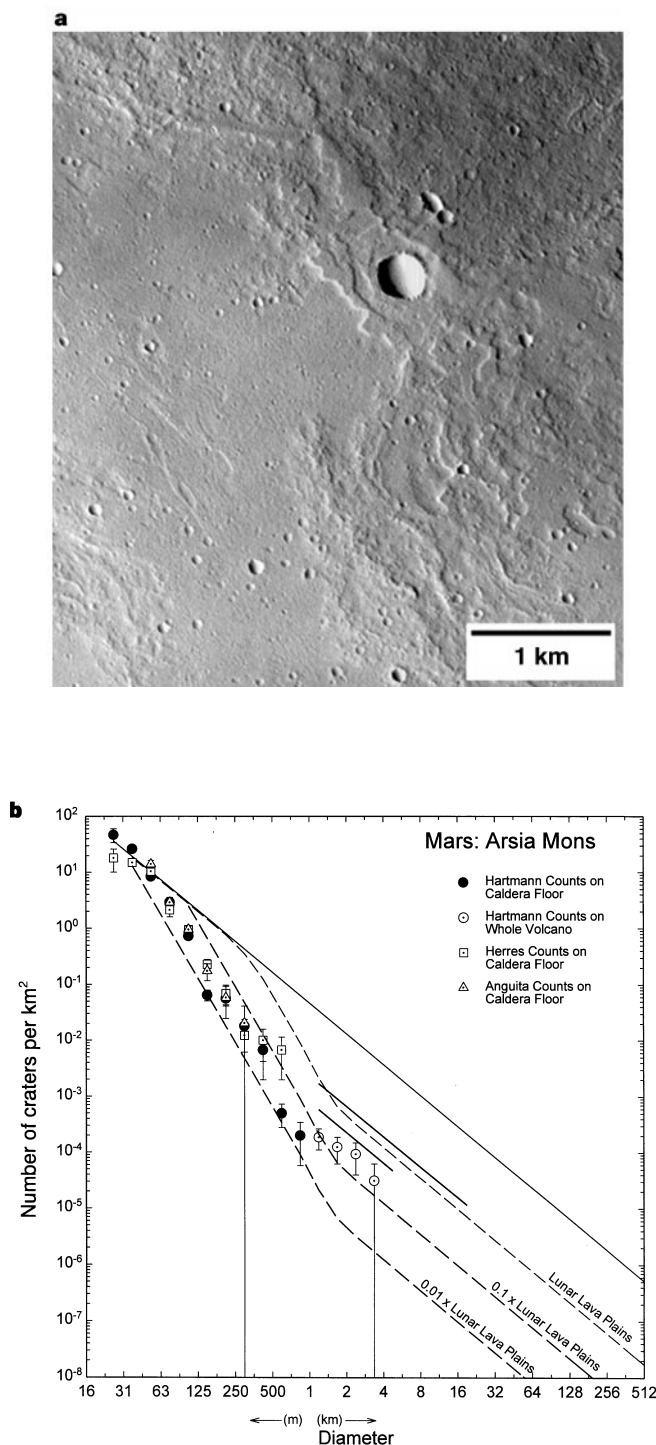


Figure 1 Crater density on Arsia Mons. **a**, Portion of MOC image no. 3308, showing portion of summit caldera floor on Arsia Mons. **b**, Crater counts for the caldera floor and flanks of Arsia Mons, superimposed on reference lines scaled to crater populations on lunar lava plains. The two short solid lines represent the stratigraphic definition of the division between three eras of Martian history: Amazonian (lower part of graph), Hesperian (between the lines) and Noachian. The counts suggest that the caldera floor is younger than the flanks of the volcano. (The crater counts were obtained from the images by more than one person, to avoid bias and to test repeatability; the name of the person is given in the key.)

distribution, and support the contention that the recent martian production function matches that of the Moon at all observed sizes. At the upper left of Fig. 1b, the curve intersects the proposed saturation equilibrium line (shown solid) at diameter $D \approx 60$ m. This behaviour is analogous to that found in the lunar maria, where the steep branch hits the saturation line at $D \approx 300$ m.

These data contain age information. The reference lines (Fig. 1b) show that the crater density in the summit caldera is only 2–10% of that found in the lunar maria; the crater density on the outer slopes may be roughly 3–10 times that. Our interpretation is that the caldera lavas are relatively young and that no substantial oblitative losses have occurred for craters down to $D = 60$ m, or depths as shallow as ~ 10 m. A review¹² of asteroid and cometary data and cratering physics suggests that the actual martian crater production rate in recent geological time is 1–4 times the mean post-mare lunar rate, with a best estimate of 2. This best estimate yields an age estimate for Arsia Mons caldera lava flows of roughly 40–200 Myr, and the outer flank ages would be several times older, with uncertainties of a factor of 2–3.

The good match between the slope of the martian and the lunar crater diameter distributions, at $60 \text{ m} < D < 1 \text{ km}$, indicates that there has not been enough dust infill in this region to remove many craters larger than 60 m. Nonetheless, although the altitude is ~ 26 km above the mean surface of Mars, we see direct evidence for some dust deposition in certain areas on the caldera rim. Figure 2 shows a portion of MOC image no. 3308 in a region of horst-graben structure just outside the north caldera rim of Arsia Mons. Rilles and other textures are clearly seen on the horst surfaces, but are muted or covered entirely by smooth deposits on the lower graben floors, especially in smooth drifts banked against the edges of the grabens. The dust source may be fallout from global dust storms that inject dust into layers as high as 35–40 km in the martian atmosphere^{15,16}.

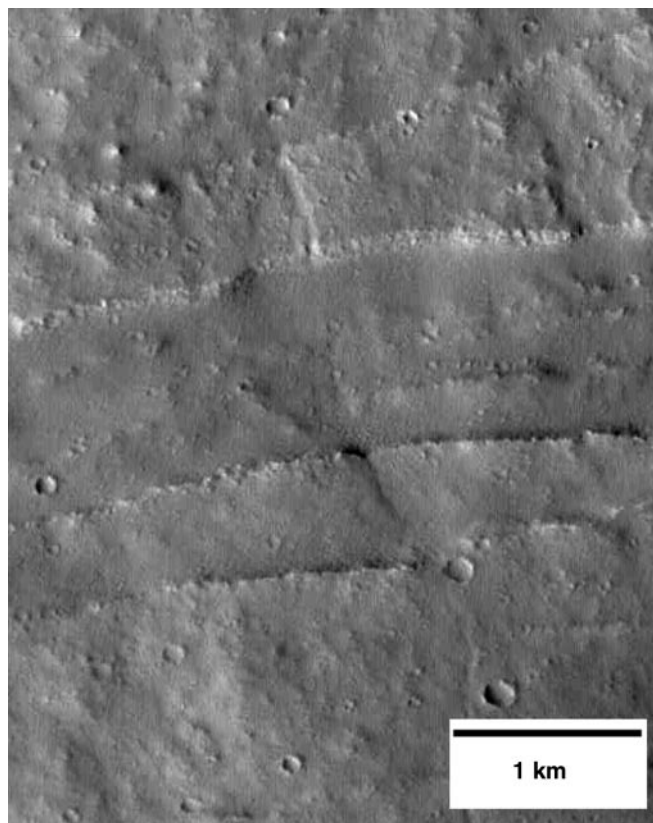


Figure 2 Horst-graben structure concentric with Arsia Mons caldera rim. Parts of graben floors show evidence of dust deposits. See text for details.

An additional issue clarified by these data involves the minimum crater size on Mars. Viking lander analysts concluded that the crater population cut off below diameter $D = 50$ m, due to atmospheric breakup of bolides¹⁷. The MOC images give the first chance to test that prediction. We find no cut-off down to $D \approx 16$ m. Many local regions are resurfaced by dust drifts and have few small craters, but other nearby areas show old surfaces where crater numbers increase smoothly as D decreases, down to 16 m and less.

The comparison between young areas and ancient upland areas

on Mars is striking. Figure 3a shows a moderately heavily cratered upland area near Nirgal Vallis in MOC image no. 605, and crater counts are given in Fig. 3b. Figure 4a shows a heavily cratered terrain in MOC image no. 2303 on the floor of the crater Schiaparelli, a large, old crater which in turn is superposed on one of the most heavily cratered martian terrains, Arabia Terra; crater counts are given in Fig. 4b. As is characteristic of all heavily cratered areas, the crater counts for both these areas show a pronounced flattening of the primary crater branch from $1 \text{ km} < D < 45 \text{ km}$. This flattening

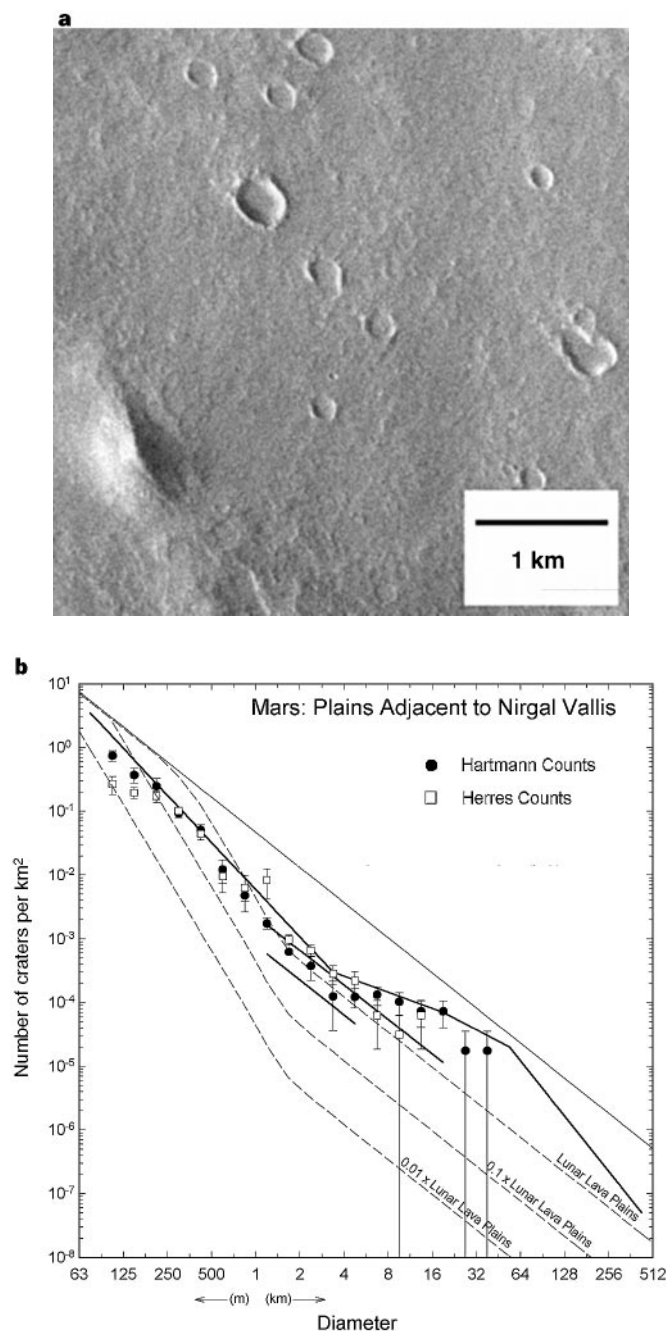


Figure 3 Crater density in the area around Nirgal Vallis. **a**, Moderate crater density in plains adjacent to Nirgal Vallis, in MOC image no. 605. We note the degraded states of some craters. **b**, As in Fig. 1b but for the older upland region around Nirgal Vallis, including area of image **a**. The solid, bent line is a calculated steady-state line showing the Öpik effect for craters with constant net dust deposition of $10^{-6} \text{ m yr}^{-1}$ (W.K.H., unpublished results). The counts suggest an old surface, more than 3 Gyr old, in which smaller craters have been lost by oblitative processes, such as dust infill.

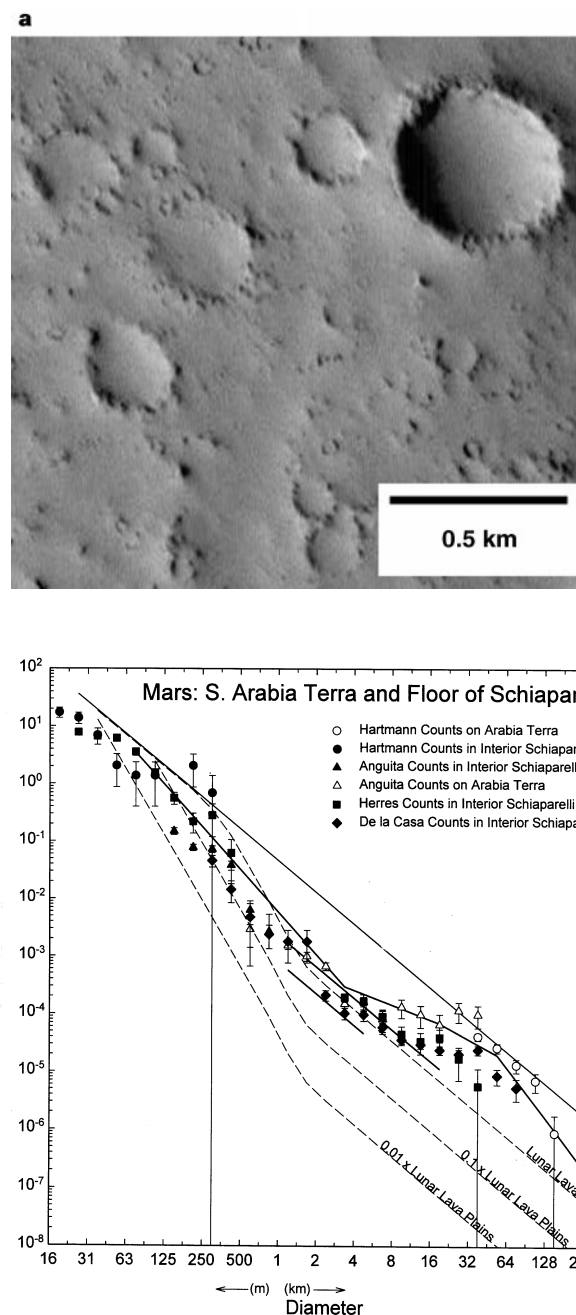


Figure 4 Crater density in the area around the crater Schiaparelli. **a**, Portion of MOC image no. 2303 showing heavily cratered portions of the floor of the crater Schiaparelli. Many craters are severely degraded. **b**, As in Fig. 1b but for the floor of crater Schiaparelli and the surrounding old region of Arabia Terra, including area of image **a**. The largest craters in Arabia Terra appear near saturation, and the surface is probably ~ 4 Gyr old. Schiaparelli appears somewhat younger, perhaps 3–4 Gyr old. Smaller craters in this region have apparently been lost by oblitative effects; the oldest visible 20-m craters may date back no more than ~ 10 Myr.

was detected by Öpik as early as 1965 and was attributed by him to deposition of material in craters, preferentially obliterating small craters¹. It has subsequently been interpreted as evidence of long-term erosion, deposition, and lava flooding of martian craters, especially in the earlier parts of Mars' history^{2–6}, although it has also been suggested that the early production function on Mars was flatter than the present function¹⁸. Our MOC data show that the steep branch of the curve in old areas also appears distinctly flatter than on the Moon; in addition, the MOC images (Figs 3a and 4a) reveal a range of degradation states among 100-m-scale craters. These states range from fresh craters to craters with dune deposits on the floor, to craters whose floors are filled and whose rims barely protrude above the dust. This fits the view that small craters have been lost by dust infill and blanketing, and the flattening of the production-function curves, at least at small diameters, is thus attributed to the Öpik effect.

The heavy, bent, solid line in Figs 3b and 4b is a predicted steady-state line for the Öpik effect infilling of craters. This curve is generally derived in refs 3 and 4, but has been modified by unpublished calculations of one of us (W.K.H.), taking into account the depth–diameter relation for fresh martian craters¹⁸. The average net deposition rate in crater floors, assumed in this curve, is $\sim 10^{-6}$ myr⁻¹, consistent with other estimates^{2–7,14}. The predicted curve is a good fit for the data. The conclusion is that on the oldest martian uplands, smaller craters are probably in a rough equilibrium with local obliteration processes, at least if we average over large enough areas. A similar statement applies to Earth, but with higher obliteration rates.

The comparison of crater size distributions on the old surfaces and young lava surfaces of Mars, and the lunar mare lava plains, indicates the wide range of surface ages on Mars, relative to the Moon; this comparison supports a conclusion that the youngest large-scale lava eruptions on Mars are much younger than on the Moon, having occurred in the last few per cent of martian time. The discovery of martian basaltic meteorites with crystallization ages of 1.3 Gyr or younger¹⁹ supports this conclusion. The crater statistics that we report here suggest that volcanism is continuing on Mars in current geological time. □

Received 16 September; accepted 14 December 1998.

- Öpik, E. J. Mariner IV and craters on Mars. *Irish Astron. J.* **7**, 92–104 (1965); The Martian surface. *Science* **153**, 255–265 (1966).
- Hartmann, W. K. Martian cratering (Paper I). *Icarus* **5**, 565–576 (1966).
- Chapman, C., Pollack, J. & Sagan, C. *An Analysis of the Mariner 4 Photography of Mars* (Spec. Rep. 268, Smithsonian Astrophys. Obs., 1968).
- Hartmann, W. K. Martian cratering III: Theory of crater obliterations. *Icarus* **15**, 410–428 (1971).
- Jones, K. L. Evidence for an episode of crater obliteration intermediate in Martian history. *J. Geophys. Res.* **79**, 3917–3931 (1974).
- Chapman, C. R. Cratering on Mars. I. Cratering and obliteration history. *Icarus* **22**, 272–291 (1974).
- Hartmann, W. K. Martian cratering, IV: Mariner 9 initial analysis of cratering chronology. *J. Geophys. Res.* **78**, 4096–4116 (1973).
- Neukum, G. & Hiller, K. Martian ages. *J. Geophys. Res.* **86**, 3097–3121 (1981).
- Strom, R. G., Croft, S. K. & Barlow, N. G. in *Mars* (ed. Kieffer, H.) 383–423 (Univ. Arizona Press, Tucson, 1992).
- Hartmann, W. K. Planetary cratering 1. Lunar highlands and tests of hypotheses on crater populations. *Meteoritics* **30**, 451–467 (1995).
- Plaut, J., Kahn, R., Guinness, E. & Arvidson, R. Accumulation of sedimentary debris in the south polar region of Mars and implications for climate history. *Icarus* **76**, 357–377 (1988).
- Hartmann, W. K. *et al.* in *Basaltic Volcanism on the Terrestrial Planets* (eds Basaltic Volcanism Study Project) 1050–1129 (Pergamon, Elmsford, NY, 1981).
- Hartmann, W. K. & Gaskell, R. W. Planetary cratering 2: Studies of saturation equilibrium. *Meteorit. Planet. Sci.* **32**, 109–121 (1996).
- Malin, M. C. *et al.* Early views of the Martian surface from the Mars Orbiter camera of Mars Global Surveyor. *Science* **279**, 1681–1685 (1998).
- Binder, A. B. *et al.* The geology of the Viking 1 lander site. *J. Geophys. Res.* **82**, 4439–4451 (1977).
- Gault, D. E. & Baldwin, B. S. Impact cratering on Mars: some effects in the atmosphere. *Eos* **51**, 343 (1970).
- Carr, M. H. & Viking Orbiter Team *Viking Orbiter View of Mars* (Spec. Publ. 441, NASA Washington DC, 1980).
- Cinalla, M. J. in *Impact and Explosion Cratering* (eds Roddy, D. J., Pepin, R. O. & Merrill, R. B.) 575–592 (Pergamon, Elmsford, NY, 1977).
- Nyquist, L. *et al.* A single-crater origin for Martian shergottites: Resolution of the age paradox? *Lunar Planet. Sci.* **29**, 1688 (1998).

Acknowledgements. We thank G. Herres, G. Esquerdo, and, in Madrid, J. Anguita and M. de las Casas, for assistance with crater counts and data processing. We also thank D. Berman and G. Hartmann for editorial assistance.

Correspondence and requests for materials should be addressed to W.K.H. (e-mail: hartmann@psi.edu).

Groundwater formation of martian valleys

Michael C. Malin* & Michael H. Carr†

*Malin Space Science Systems, PO Box 910148, San Diego, California 92191-0148, USA

†US Geological Survey, 345 Middlefield Road, Menlo Park, California 94025, USA

The martian surface shows large outflow channels, widely accepted as having been formed by gigantic floods that could have occurred under climatic conditions like those seen today^{1–5}. Also present are branching valley networks that commonly have tributaries^{1–8}. These valleys are much smaller than the outflow channels and their origins and ages have been controversial. For example, they might have formed through slow erosion by water running across the surface, either early or late in Mars' history^{9–13}, possibly protected from harsh conditions by ice cover^{14–16}. Alternatively, they might have formed through groundwater or ground-ice processes that undermine the surface and cause collapse, again either early or late in Mars' history^{3,4}. Long-duration surface runoff would imply climatic conditions quite different from the present environment. Here we present high-resolution images of martian valleys that support the view that ground water played an important role in their formation, although we are unable as yet to establish when this occurred.

Images acquired by the Mars Orbiter Camera (MOC) during the aerobraking phase (September 1997 to February 1998) of the Mars Global Surveyor mission typically have resolutions in the range 4–8 m per pixel, in most cases a factor of 20–50 times better than previous imaging^{17,18}. The images reveal new details about the valleys that strongly support an origin by fluid erosion. Although apparent drainage networks are observed locally, dissection of the adjacent upland surface, as might be expected if the fluid had an atmospheric rather than a subsurface source, is not seen. The lack of eroded uplands adjacent to martian fluvial valleys and the implication for a localized water source was previously noted in studies of Viking Orbiter images^{19,20}.

Figure 1 is an image of Nanedi Vallis—an 800-km-long valley that appears incised into cratered plains north of the Valles Marineris. Nanedi Vallis is one of the longest and freshest-appearing of the martian valley networks. Despite its length it has only a few short tributaries, and no obvious catchment area. It starts close to the equator at 49°W in an area where there is other evidence for groundwater action, including the source of the outflow channel, Shalbatana Vallis. The circuitous path of the valley seen here appears to have been inherited from surficial fluid movement, although the source of the fluid is not apparent. Such arcuate and reversing paths are difficult, if not impossible, to create by headward erosion (that is, progressively upstream towards the source) of a stream, lending support to an interpretation, based on visual appearance, that the valley formed by entrenchment of an originally meandering flow. This interpretation is further strengthened by the observation of an interior channel, presumably the specific course of the valley-forming fluid. However, as with many entrenched valley systems on Earth, mass movements accompanying groundwater action are likely to have created much of the relief and width of the presently observed valley. It could be argued that valley formation reflected groundwater processes fed by precipitation, and that the lack of dissection of the adjacent plain is the result of high permeability of the near-surface materials²¹. However, the total absence of metre-scale dissection here and elsewhere on Mars, and the near-absence of upstream tributaries (suggesting a spatially limited source), support a subsurface rather than atmospheric source.

These MOC images show clearly that the uncratered and often



Figure 1 Portion of MOC image no. 8704, showing Nanedi Vallis (5.5° N, 48.4° W), a "run-off channel"^{3,4} north of Valles Marineris and south of Chryse Planitia. This valley shows a sinuous path strongly suggestive of meanders (topmost portion of image), large benches indicating earlier floor levels (upper third of image centre), and a central channel (also upper third, right side). These features support the idea that this valley was formed by sustained fluid flow. Viewed somewhat obliquely (emission or viewing angle, 54.9°) and at relatively high Sun angle (incidence angle, 38.1°), this image has a downtrack scale of 14.5 m per pixel and a crosstrack scale of 9.5 m per pixel. (The inset in each figure shows the outline of an MOC image as a white box overlain on a lower-resolution Viking Orbiter image.)

flat floors of the larger valleys^{22,23} are at least partly, if not mainly, the result of aeolian infill accumulated in the valleys after they formed. Various stages of fill are observed. Dune fields cover the flat valley floors of Nirgal Vallis (MOC image no. 0605), Samara Vallis (no. 3004) and an unnamed valley at 2.8° S, 234.3° W (no. 3606). In Evros Vallis (12.9° S, 343.9° W, no. 10604) the dunes cover only a narrow swath, roughly 200 m across, between the talus slopes from the opposing valley walls. In Nanedi Vallis (Fig. 1, no. 8704) large dunes are absent, small dunes are seen only on some portions of the floor, and the opposing talus slopes often meet at the centre of the valley, covering the central channel. The presence of the aeolian fill may explain why channels have been so rarely observed within the valleys: the flat valley floors are not flood plains, but rather the consequence of later infilling.

One of the most striking attributes of the surface of Mars as seen in MOC images is the absence of dissection, despite their 4–8 m per pixel resolution, and the clear visibility of other landforms (for example, impact craters, layered rock outcrops, and fault scarps) that indicate where mantles or debris blankets do or do not occur. This is especially true adjacent to martian valleys: where tributaries are present they tend to be short, nearly as broad as the main valley, and to have discordant junctions. Many valleys appear to start abruptly, almost full size, with no small tributaries. Patterns of rimless, surface pits suggest that removal of subsurface materials contributes to valley formation (Fig. 2). These attributes further suggest that (1) although some valleys formed at least in part by sustained erosion by surficial flow, such flow was neither ubiquitous nor contributory, as would be expected for an atmospheric source,

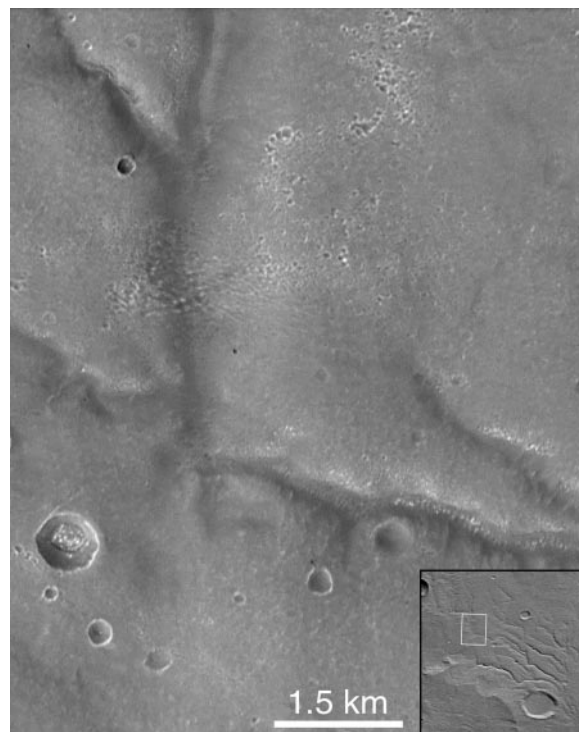


Figure 2 Portion of Corasis Fossae valleys (MOC image no. 8205). These subdued valleys (35.6° S, 75.4° W) show structural control, as do pits in the centre of the upper half of the image. The pattern of pitting suggests that removal of subsurface support may have been important in valley formation. The adjacent upland surface is not dissected. This slightly oblique view (emission, 20.7°) was taken at low incidence angle (25.4°); the downtrack scale is 11.5 m per pixel and the crosstrack scale is 6.6 m per pixel.

and (2) valley growth occurred both by widening by mass wasting and deepening by incision and/or removal of subsurface support.

A different valley system (unique in MOC observations) is seen within the 145-km-diameter crater Bakhuisen (Fig. 3). There, small valleys, typically 0.5 km wide, with numerous tributaries, occur only on the interior wall of the crater. The valleys extend from below the rim crest, down the crater wall and terminate against the flat floor of the crater (Viking image no. 618A23). The Bakhuisen networks are substantially denser and better integrated than any seen elsewhere in MOC data, but an origin related to precipitation appears unlikely, as no valleys are visible outside the rim crest. Surficial runoff is suggested, but from a restricted source within the crater: groundwater seepage from beds exposed within the crater wall, possibly aided by hydrothermal activity associated with the heat produced by the impact cratering process^{24,25}, seems the most reasonable model.

In the most dissected terrains, such as those seen near the upper reaches of Parana Vallis (Fig. 4), the fluid responsible for the deterioration of the landscape is not limited to a small number of discrete sources. Rather, the appearance of many short, adjacent, tapered 'gullies' contributing to a broad, shallow irregularly-outlined trunk suggests a pervasive source. Even in this case, however, there is no contributing network of smaller channels on the upland surface to suggest an atmospheric source, and the drainage pattern, such as it is, is immature—it shows large differences in size between reaches of sequential order and the pattern shows little or no topographic, structural, or material control, yet can hardly be termed 'dendritic'. Valleys such as these argue further for subsurface fluid sources.

The range in ages of the martian valleys seen in MOC images has

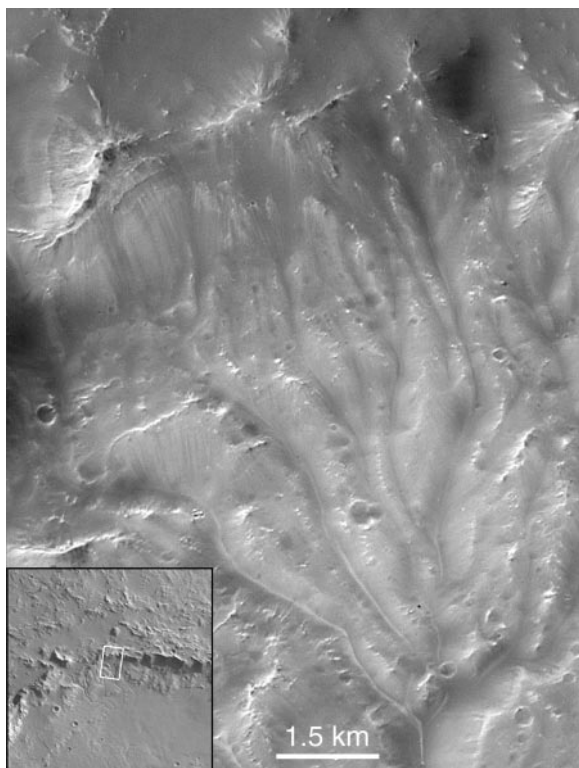


Figure 3 Portion of channels on the wall of Bakhuyzen crater (MOC image no. 10605). These channels (22.1°S, 344.9°W) are the best examples of integrated drainage reminiscent of terrestrial systems. The pattern is topographically controlled; the relationships are emphasized by light-coloured sediments viewed in this low-incidence-angle (11.2°), nadir viewing (emission, 1.5°) image. The crater rim is marked by the escarpment running diagonally from the middle left to the upper right of the image (downtrack scale, 8.4 m per pixel; crosstrack scale, 5.8 m per pixel). No channels are seen outside the crater rim. This suggests that the source of the fluid was confined within the crater.

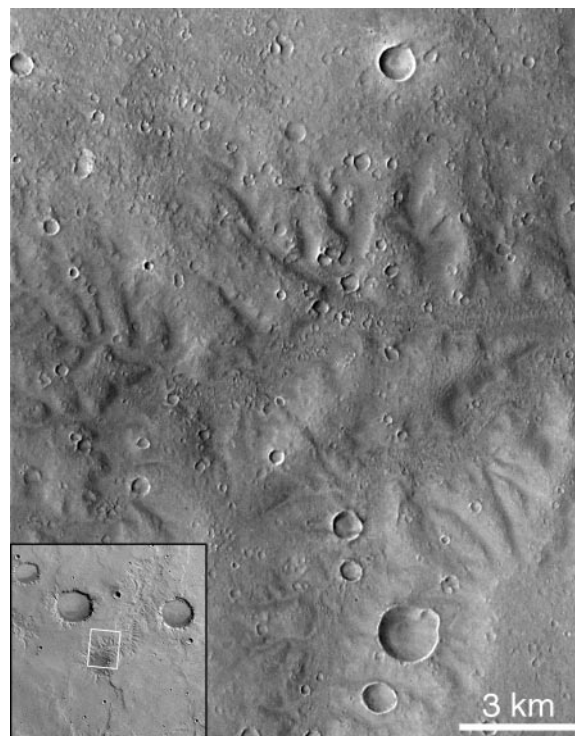


Figure 4 Portion of the dissected terrain southeast of Parana Valles (MOC image no. 7705). This heavily gullied landscape (25.9°S, 8.3°W) shows the highest "drainage density" yet seen in MOC images. This image is somewhat lower in resolution (downtrack scale, 21.4 m per pixel; crosstrack scale, 14.3 m per pixel), but in other parameters is comparable to Figs 1 and 2 (incidence angle, 27.5°, emission, 14.5°).

not yet been determined. Comparison of the abundance of impact craters in Figs 1 and 4 might lead one to conclude a great disparity in age. However, as noted over 20 years ago²⁶, the similar number of craters in the diameter range 4–10 km despite a large difference in crater abundances at larger diameters strongly suggests that these surfaces, and the valleys found on them, formed during the highly nonlinear portion of the bombardment history.

The images being acquired by the Mars Orbiter Camera support an origin of the valley networks by fluid erosion; the nature of the landforms suggest that this fluid was water. In most cases the source appears to have been ground water. A problem still arises as to how the groundwater system was charged so that it could feed the streams. Early warm climates with an active hydrologic cycle have been proposed. The MOC data provide no substantive evidence for such early climates. Rather, they show that surface features—even those in the most heavily cratered terrains—reflect processes that have acted to obscure all traces of these early climates, or that have proceeded without regard to environmental conditions. Although this is not, and should not be taken as, an argument against such early climates, it stands as a substantial warning that future spacecraft (such as landers or sample-return vehicles) may not be able to gain access to materials that reflect these early times. □

Received 23 September; accepted 24 December 1998.

1. Baker, V. R. *The Channels and Valleys of Mars* (Univ. Texas Press, Austin, 1982).
2. Carr, M. H. *Water on Mars* (Oxford Univ. Press, 1992).
3. Sharp, R. P. & Malin, M. C. Channels on Mars. *Geol. Soc. Am. Bull.* **86**, 593–609 (1975).
4. Mars Channel Working Group. Channels and valleys on Mars. *Geol. Soc. Am. Bull.* **94**, 1035–1054 (1983).
5. Baker, V. R. & Milton, D. J. Erosion by catastrophic floods on Mars and Earth. *Icarus* **23**, 27–41 (1974).
6. Baker, V. R. *et al.* Spring sapping and valley network development. *Geol. Soc. Am. Spec. Pap.* **252**, 235–265 (1990).

7. Carr, M. H. Formation of martian flood features by release of water from confined aquifers. *J. Geophys. Res.* **84**, 2995–3007 (1979).
8. Pieri, D. C. Martian valleys: Morphology, distribution and age. *Science* **210**, 895–897 (1980).
9. Baker, V. R. *et al.* Ancient oceans, ice sheets and the hydrologic cycle on Mars. *Nature* **332**, 589–594 (1991).
10. Masursky, H., Boyce, J. M., Dial, A. L., Schaber, G. G. & Strobel, M. E. Classification and time of formation of martian channels based on Viking data. *J. Geophys. Res.* **82**, 4016–4038 (1977).
11. Pollack, J. B. Climate change on the terrestrial planets. *Icarus* **37**, 357–377 (1979).
12. Craddock, R. A. & Maxwell, T. A. Geomorphic evolution of the martian highlands through ancient fluvial processes. *J. Geophys. Res.* **98**, 3453–3468 (1993).
13. Gulick, V. C. & Baker, V. R. Fluvial valleys and martian palaeoclimates. *Nature* **341**, 514–516 (1989).
14. Wallace, D. & Sagan, C. Evaporation of ice in planetary atmospheres: Ice-covered rivers on Mars. *Icarus* **39**, 385–400 (1979).
15. Carr, M. H. The stability of streams and lakes on Mars. *Icarus* **56**, 476–495 (1983).
16. Squyres, S. W. & Kasting, J. F. Early Mars: How warm and how wet? *Science* **265**, 744–748 (1994).
17. Malin, M. C., Danielson, G. E., Ravine, M. A. & Soulanille, T. A. Design and development of the Mars Observer Camera. *Int. J. Imaging Syst. Technol.* **3**, 76–91 (1991).
18. Malin, M. C. *et al.* Early views of the martian surface from the Mars Orbiter Camera of Mars Global Surveyor. *Science* **179**, 1681–1685 (1998).
19. Gulick, V. C., Tyler, D., McKay, C. P. & Haberle, R. M. Episodic ocean-induced CO₂ greenhouse on Mars: Implications for fluvial valley formation. *Icarus* **130**, 68–86 (1997).
20. Gulick, V. C. Magmatic intrusions and a hydrothermal origin for fluvial valleys on Mars. *J. Geophys. Res.* **103**, 19365–19387 (1998).
21. Baker, V. R. & Partridge, J. B. Small Martian valleys: pristine and degraded morphology. *J. Geophys. Res.* **91**, 3561–3572 (1986).
22. Goldspiel, J. M., Squyres, S. W. & Jankowski, D. G. Topography of small martian valleys. *Icarus* **105**, 479–500 (1993).
23. Carr, M. H. The martian drainage system and the origin of valley networks and fretted channels. *J. Geophys. Res.* **100**, 7479–7507 (1995).
24. Brackenridge, G. R., Newsom, H. E. & Baker, V. R. Ancient hot springs on Mars: Origin and paleosignificance of small martian valleys. *Geology* **13**, 859–862 (1985).
25. Newsom, H. E. Hydrothermal alteration of impact melt sheets with implications for Mars. *Icarus* **44**, 207–216 (1980).
26. Soderblom, L. A., Condit, C. D., West, R. A., Herman, B. M. & Kreidler, T. J. Martian planetwide crater distributions: Implications for geologic history and surface processes. *Icarus* **22**, 239–263 (1974).

Acknowledgements. We thank J. Warren, R. Adair and M. Caplinger for efforts in support of aerobraking operations, and the Mars Surveyor Operations Project and its personnel for enabling aerobraking science observations. M.C.M. was supported by JPL.

Correspondence and requests for materials should be addressed to M.C.M. (e-mail: malin@mss.com).

Bright dunes on Mars

P. C. Thomas*, M. C. Malin†, M. H. Carr‡, G. E. Danielson§, M. E. Davies||, W. K. Hartmann¶, A. P. Ingersoll§, P. B. James#, A. S. McEwen*, L. A. Soderblom** & J. Veveřka*

* Center for Radiophysics and Space Research, Cornell University, Ithaca, New York 14853, USA

† Malin Space Science Systems, PO Box 910148, San Diego, California 92191, USA

‡ US Geological Survey, Menlo Park, California 94025, USA

§ Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA

|| The Rand Corporation, Santa Monica, California 90406, USA

¶ Planetary Science Institute, Tucson, Arizona 85719, USA

Department of Physics and Astronomy, University of Toledo, Toledo, Ohio 43606, USA

* Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 86001, USA

** US Geological Survey, Flagstaff, Arizona 86001, USA

Seasonal changes observed on the surface of Mars can in part be attributed to the transport of geological materials by wind¹. Images obtained by orbiting spacecraft in the 1970s showed large wind-formed features such as dunes, and revealed regional time-varying albedos that could be attributed to the effects of dust erosion and deposition. But the resolution of these images was insufficient to identify different types and sources of aeolian materials, nor could they reveal aeolian deposits other than large dunes or extensive surface coverings that were redistributed by dust storms. Here we present images of Mars with up to 50 times better resolution. These images show that martian dunes include at least two distinct components, the brighter of which we interpret to be composed of relatively soft minerals, possibly sulphates. We also find large areas of the martian surface that have several metres or more of aeolian mantle lacking obvious bedforms.

Large dunes and dune fields on Mars are some of the darkest materials on the planet^{2,3}. The dark dunes have been chiefly interpreted to be silicate minerals high in iron and magnesium (mafic or ultramafic) or dark lithic fragments^{4–6}. Viking Orbiter images showed a few instances of poorly resolved brighter dunes⁷ as well as dunes indistinguishable from average Mars albedo, many of which were interpreted to be partially dust-covered and inactive^{8,9}. The images from the Mars Orbiter Camera (MOC) on the Mars Global Surveyor spacecraft reveal several instances of brighter dunes that have distinctive morphologies (Fig. 1). The characteristics of the bright dunes include: (1) transverse bedforms; (2) small sizes, with maximum lengths (parallel to inferred wind direction) of ~40 m, typically 15–20 m, and width-to-length ratios as high as 15, typically over 6; (3) occurrence near outcroppings of bright materials; (4) estimated albedos at visible wavelengths as high as 0.32, or about twice that of the materials in large, dark dune fields.

From the extensive global imaging of the Viking Orbiters, we know that, unlike the dark dunes, the bright material does not form regional dune fields^{10,11}. The few hundred MOC high-resolution images provide only a preliminary view of the distribution of small features, but in those images returned so far the bright dunes are within a few kilometres of a variety of outcroppings of bright materials. The apparent restriction of the bright dunes to small distances from outcroppings of source materials suggests that their materials are not physically resistant to saltation impacts as are quartz and feldspars. Because sulphur is known to be abundant in Martian soils, probably in oxidized, sulphate forms^{12,13}, we speculate that the bright dunes are composed of gypsum, or some other sulphate. Soft materials such as gypsum would have limited life-

times travelling in saltation on Mars because the high velocity of saltating grains in the thin martian atmosphere¹⁴ fosters rapid breakdown to smaller particles that can be carried in suspension. Sulphates on Mars could come directly from playa deposits, or other concentrations derived from ground water, flood or hydrothermal activity¹⁵. If the martian bright dunes involve sulphates, they could either be separate sulphate particles or even other mineral particles cemented by sulphates. On Earth, substantial sulphate dunes are found immediately adjacent to sources such as playa deposits, and change character rapidly downwind^{16,17}. In contrast, the dark dunes on Mars are able to exist far from obvious sources, though local sources for some intracrater dark dunes have been suggested^{18,19}.

Are the bright bedforms currently active? In seven of the ten unambiguous instances, the effective wind direction inferred from bright dunes differs by less than 15° from the local wind streak orientations¹⁸; the other three are different by 50–90°, but are clearly affected by local topography. Because the bright dunes are at low latitudes within the global Hadley circulation, the strongest winds would be expected to change direction substantially (by 90° or more) over the 51,000-year cycle of season of perihelion^{20,21}. Thus the bright dune orientations are consistent with geologically recent winds. The image scales (4–8 m per pixel) allow estimates of freshness only for the timescales needed to modify features several metres in size, which (because of uncertainties in particle properties, wind velocities, and the highly nonlinear relation of transport flux to wind shear stress) could be from a few years to thousands of years^{22,23}. The longer timescales may stretch the definition of 'currently active', especially if they are larger than half of the 51,000-year cycle of perihelion season. With one exception the bright dunes show no significant superposed features, obvious erosional gaps or other evidence of degradation. The exception (Fig. 1b) has dark, barchanoid (that is, crescent-shaped) forms and sheets that postdate remnants of bright bedforms that have slightly different orientations. In this instance, the bright bedforms might be left over from a time of different wind patterns or sediment availability; the absolute timescale for the change is unknown.

The retention of albedo contrasts on Mars over time requires surface cleaning to keep them free from the ubiquitous deposition of atmospheric dust¹. This cleaning is usually thought to depend on saltation-initiated activity^{1,18}. The albedo contrasts in Fig. 1 indicate that even if the bright bedforms are not at present active, the surrounding darker materials have been cleaned in the last Mars year by the wind^{5,24}.

The shapes and arrangements of most of the dunes imaged by the MOC suggest formation transverse to the effective winds. This observation accords with previous findings that dunes attributable to highly variable sand-transporting winds are rare on Mars. Other indicators of winds, as well as general circulation models, suggest a predominance of nearly unimodal direction of sand transport on Mars^{25,26}. (The strong directionality of Mars' winds derives in large part from the orbital eccentricity and rapid thermal response of the thin atmosphere²⁵.) The MOC data are replete with small, transverse dunes which show that, even on small spatial scales, sand-transporting winds are strongest in one primary direction. One of the striking features of the MOC images is the small size of the few dunes formed by multiply effective winds. Unlike the Earth, where reversing wind regimes can form very large star dunes, Mars shows no instances where reversing winds have acted to trap large quantities of sand in individual dunes.

A large fraction of the martian surface has been suspected to be covered by aeolian mantles which might consist of dust, dunes smaller than Viking resolutions, or a variety of sand-sheet-like deposits^{27,28}. The MOC data show that while small dunes are common, aeolian mantles free of visible bedforms cover large areas on Mars. Craters less than a few hundred metres in diameter show filling by materials having a variety of albedos and surface morphologies; this fill affects crater density statistics²⁹ (Fig. 2). Some

filling is smooth, and nearly continuous with the surroundings, leaving only remnants of crater rims exposed. These observations suggest depths of fill averaging nearly 0.1 of the crater's diameter (assuming normal fresh-crater shape); in some areas, craters up to 600 m across have average accumulations of ~ 50 m (ref. 29).

Likely rates of aeolian dust deposition on Mars are very uncertain, though present atmospheric loading has the potential for nearly $10^{-3} \text{ g cm}^{-2} \text{ yr}^{-1}$ (refs 20, 21); this value would be slightly low for modern continental dust deposition on Earth, but high for ocean aeolian deposition rates³⁰. Such rates could fill a 100-m diameter

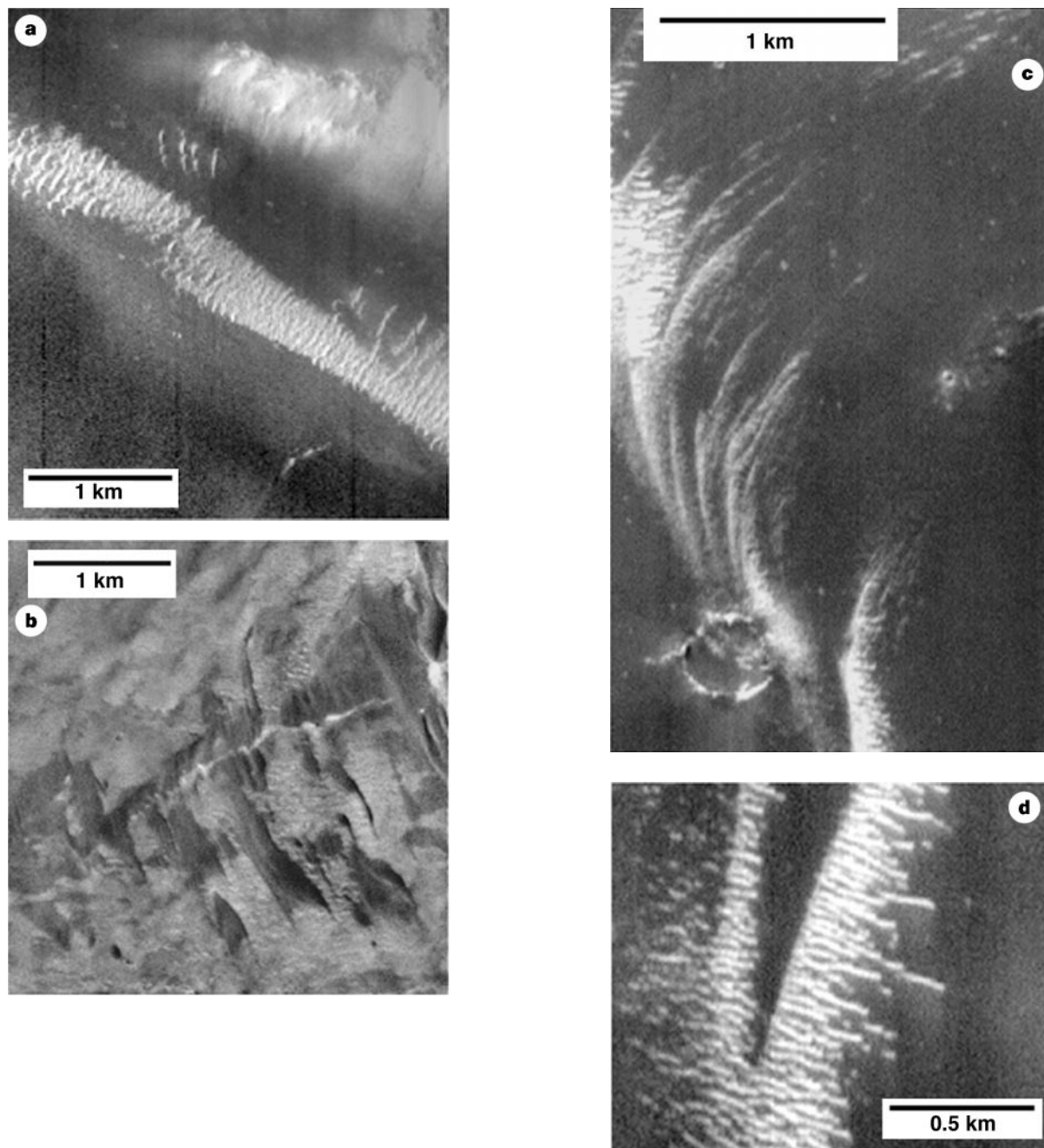


Figure 1 MOC images of bright and dark dunes on Mars. Portions of individual frames are shown, and have been moderately contrast-enhanced from the raw data; north is approximately up in all images. **a**, Image no. 2306 (orbit 23, image 06), centred near 5°S , 341°W . Original image scales were 4.5 by 7.9 m per pixel; all images rectified here to square pixels. The bright dunes show slightly crescentic forms made by winds blowing towards the southeast; some of the areas between dune ridges are also covered by relatively bright materials. The top, right-hand part of the image shows bright ridge outcrops; these occur in other nearby areas in the full image and in other nearby images. **b**, Image no. 4206, image scales 4.1 by 6.2 m per pixel, 6.7°S , 72.5°W , showing dark sheets, barchans, and degraded, smaller bright dunes. The dark dunes indicate winds to the south-southeast; the brighter dunes probably were formed by winds with net transport direction slightly more towards the south. The dark dunes appear to overlie some of the

bright ones, and thus post date the bright dunes. **c**, Image no. 11105, image scales 5.3 by 8.0 m per pixel, showing bright dunes near 7°S , 344°W in the Schiaparelli basin. This image shows the bright dunes, bright outcroppings such as crater rims, and a smooth darker surface, perhaps an aeolian mantle of substantially different materials. The bright dunes show forms probably formed transverse to winds flowing almost to the south (expected in current climate). These forms are grouped into larger curving patterns that may either be relicts from larger transverse forms in different wind regimes, or caused by subtle topographic troughs in the smooth, darker materials. **d**, Detail of part of image no. 11105, just south of the section shown in **c**. This view emphasizes the large width-to-length ratio of these bedforms, and the presence of some bright material blown off the dunes towards the south.

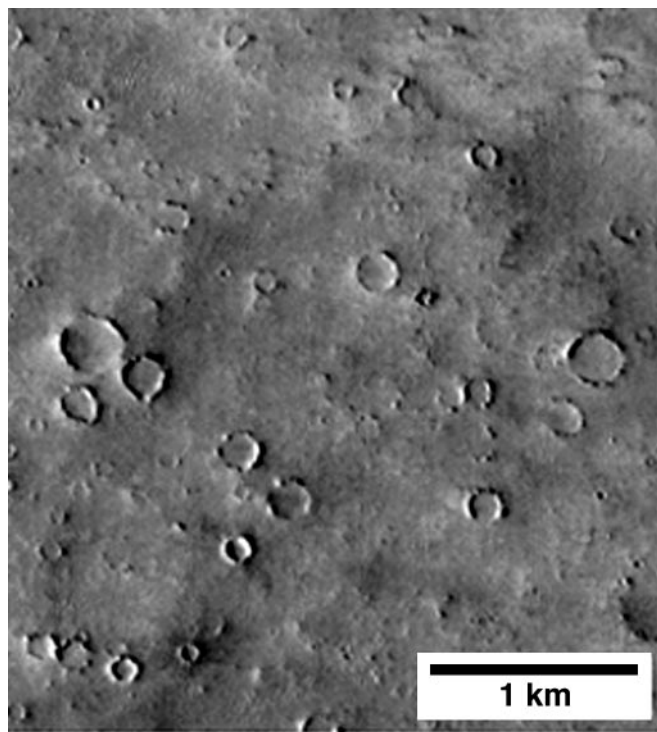


Figure 2 MOC image showing blanketing material filling several craters that are 200–400 m in diameter. Image scale, 4.1 by 7.3 m per pixel; 3.6° S, 340.5° W. The crater rims are moderately well preserved and visible, indicating that aeolian material has not simply been draped over the topography but has preferentially filled the lowest areas. The average depth of fresh 200-m craters is probably ~20 m. Rim heights above surroundings are typically much less than crater depths; thus the fill inside these craters is probably deeper than in much of the intervening plains.

crater in 10^6 yr, but the clear lesson from the MOC data is that the blankets are not uniformly distributed airfall but represent a net sum of airfall, erosion, saltation cleaning and deposition. Other small craters have fill with periodic bedforms, and erosional stripping of layers and jointed sediments. The ubiquity of martian aeolian material is emphasized by the filling of craters and structural features high on Arsia Mons, more than 20 km above the reference 6.1-mbar level in the atmosphere; these deposits indicate transport of considerable amounts of material at pressures under 1 mbar (ref. 29; unless there has been a geologically recent period of higher atmospheric pressure on Mars, which is generally considered very unlikely²¹).

The indications of distinctive materials, possibly sulphates, moved in saltation on Mars expands the known complexity of the martian sedimentary system, and implies that remnants of ancient sedimentary and weathering processes have not been completely homogenized and lost from remote or *in situ* studies. This variety may assist some future sampling efforts, but the presence of widespread mantling reduces the areas within which bedrock or sorted remnants are available for sampling. □

Received 16 September; accepted 21 December 1998.

1. Sagan, C. *et al.* Variable features on Mars, 2, Mariner 9. Global results. *J. Geophys. Res.* **78**, 4163–4196 (1973).
2. Bell, J. F., Thomas, P. C., Wolff, M. J., Lee, S. W. & James, P. B. Mineralogy of the Martian north polar sand sea from 1995 Hubble Space Telescope near-IR observations. *Proc. Lunar Planet. Sci. Conf.* **27**, 87–88 (1997).
3. Thomas, P. C. & Weitz, C. Sand dune materials and polar layered deposits on Mars. *Icarus* **81**, 185–215 (1989).
4. Edgett, K. S. & Lancaster, N. Volcaniclastic aeolian dunes: Terrestrial examples and application to Martian sands. *J. Arid Environ.* **25**, 271–297 (1993).
5. Christensen, P. R. Regional dust deposits on Mars: Physical properties, age, and history. *J. Geophys. Res.* **91**, 3533–3545 (1986).

6. Mustard, J. *et al.* The surface of Syrtis major: Composition of the volcanic substrate and mixing with altered dust and soil. *J. Geophys. Res.* **98**, 3387–3400 (1993).
7. Edgett, K. S. & Parker, T. J. “Bright” aeolian dunes on Mars: Viking orbiter observations. *Proc. Lunar Planet. Sci. Conf.* **29** (CD-ROM) (1998).
8. Ward, A. W. *et al.* Global map of aeolian features of Mars. *J. Geophys. Res.* **90**, 2038–2056 (1985).
9. Edgett, K. S. Aeolian dunes as evidence for explosive volcanism in the Tharsis region of Mars. *Icarus* **130**, 96–114 (1997).
10. Tsoar, H., Greeley, R. & Peterfreund, A. R. Mars: the north polar sand sea and related wind patterns. *J. Geophys. Res.* **84**, 8167–8182 (1979).
11. Greeley, R. & Williams, S. H. Dust deposits on Mars: The “parna” analog. *Icarus* **110**, 165–177 (1994).
12. Toulmin, P. *et al.* Geochemical and mineralogical interpretation of the Viking inorganic chemical results. *J. Geophys. Res.* **82**, 4625–4634 (1977).
13. Rieder, R. *et al.* The chemical composition of Martian soil and rocks returned by the mobile alpha proton X-ray spectrometer: Preliminary results from the X-ray mode. *Science* **278**, 1771–1774 (1997).
14. White, B. R. Soil transport by winds on Mars. *J. Geophys. Res.* **84**, 4643–4651 (1979).
15. Warren, P. H. Petrologic evidence for low-temperature, possibly flood evaporitic origin of carbonates in the ALH84001 meteorite. *J. Geophys. Res.* **103**, 16759–16773 (1998).
16. McKee, E. D. Structures of dune at White Sands National Monument, New Mexico. *Sedimentology* **7**, 3–69 (1966).
17. Jones, D. J. Gypsum–oolite dunes, Great Salt Lake desert, Utah. *Bull. Am. Assoc. Petrol. Geol.* **37**, 2530–2538 (1938).
18. Thomas, P. C., Veverka, J., Gineris, D. & Wong, L. ‘Dust’ streaks on Mars. *Icarus* **49**, 398–415 (1984).
19. Malin, M. C. *et al.* Early views of the Martian surface from the Mars Orbiter Camera of Mars Global Surveyor. *Science* **279**, 1681–1685 (1998).
20. Arvidson, R. E., Guinness, E. A. & Lee, S. Differential aeolian redistribution rates on Mars. *Nature* **278**, 533–535 (1979).
21. Kieffer, H. H. & Zent, A. P. in *Mars* (eds Kieffer, H., Jakosky, B., Snyder, C. & Matthews, M.) 1180–1220 (Univ. Arizona Press, Tucson, 1992).
22. Thomas, P. C. Present wind activity on Mars: Relation to large latitudinally zoned sediment deposits. *J. Geophys. Res.* **87**, 9999–10008 (1982).
23. Greeley, R., Skyeck, A. & Pollack, J. B. Martian aeolian processes and deposits: Comparisons with general circulation model results. *J. Geophys. Res.* **98**, 3183–3196 (1993).
24. Wells, E. N., Veverka, J. & Thomas, P. Mars: Experimental study of albedo changes caused by dust fallout. *Icarus* **58**, 331–338 (1984).
25. Lee, P. L. & Thomas, P. C. Longitudinal dunes on Mars: Relation to current wind regimes. *J. Geophys. Res.* **100**, 5381–5395 (1995).
26. Edgett, K. S. & Blumberg, D. G. Star and linear dunes on Mars. *Icarus* **112**, 448–464 (1994).
27. Zimbelman, J. R. Spatial resolution and the interpretation of Martian morphology: Implications for subsurface volatiles. *Icarus* **71**, 257–267 (1987).
28. Breed, C. S., McCauley, J. F. & Davis, P. A. Ripple blankets: geomorphic evidence for regional sand sheet deposits on Mars. *Proc. Lunar Planet. Sci. Conf.* **18**, 127 (1987).
29. Hartmann, W. K. *et al.* Evidence for recent volcanism on Mars from crater counts. *Nature* **397**, 586–589 (1999).
30. Pye, K. *Aeolian Dust and Dust Deposits* (Academic, New York, 1987).

Acknowledgements. We acknowledge the technical help of M. Caplinger, J. Warren, M. Ravine, M. Ryan, M. Ockert-Bell, I. Dauber, R. Sullivan, K. Edgett, B. Carcich, A. Fox, A. Lowenkron & M. Roth.

Correspondence and requests for materials should be addressed to P.C.T. (e-mail: thomas@cusipf.tn.cornell.edu).

Light speed reduction to 17 metres per second in an ultracold atomic gas

Lene Vestergaard Hau^{*†}, S. E. Harris[‡], Zachary Dutton^{*†} & Cyrus H. Behroozi^{*§}

^{*} Rowland Institute for Science, 100 Edwin H. Land Boulevard, Cambridge, Massachusetts 02142, USA

[†] Department of Physics, [§] Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

[‡] Edward L. Ginzton Laboratory, Stanford University, Stanford, California 94305, USA

Techniques that use quantum interference effects are being actively investigated to manipulate the optical properties of quantum systems¹. One such example is electromagnetically induced transparency, a quantum effect that permits the propagation of light pulses through an otherwise opaque medium^{2–5}. Here we report an experimental demonstration of electromagnetically induced transparency in an ultracold gas of sodium atoms, in which the optical pulses propagate at twenty million times slower than the speed of light in a vacuum. The gas is cooled to nanokelvin temperatures by laser and evaporative cooling^{6–10}. The quantum interference controlling the optical properties of the medium is set up by a ‘coupling’ laser beam propagating at a right angle to the pulsed ‘probe’ beam. At nanokelvin temperatures, the variation of refractive index with probe frequency can be made very steep. In conjunction with the high atomic density,

this results in the exceptionally low light speeds observed. By cooling the cloud below the transition temperature for Bose–Einstein condensation^{11–13} (causing a macroscopic population of alkali atoms in the quantum ground state of the confining potential), we observe even lower pulse propagation velocities (17 m s^{-1}) owing to the increased atom density. We report an inferred nonlinear refractive index of $0.18 \text{ cm}^2 \text{ W}^{-1}$ and find that the system shows exceptionally large optical nonlinearities, which are of potential fundamental and technological interest for quantum optics.

The experiment is performed with a gas of sodium atoms cooled to nanokelvin temperatures. Our atom cooling set-up is described in some detail in ref. 14. Atoms emitted from a ‘candlestick’ atomic beam source¹⁵ are decelerated in a Zeeman slower and loaded into a magneto-optical trap. In a few seconds we collect a cloud of 10^{10} atoms at a temperature of 1 mK and a density of $6 \times 10^{11} \text{ cm}^{-3}$. The atoms are then polarization gradient cooled for a few milliseconds to 50 μK and optically pumped into the $F = 1$ ground state with an equal population of the three magnetic sublevels. We then turn all laser beams off and confine the atoms magnetically in the ‘4 Dee’ trap¹⁴. Only atoms in the $M_F = -1$ state, with magnetic dipole moments directed opposite to the magnetic field direction (picked as the quantization axis), are trapped in the asymmetric harmonic trapping potential. This magnetic filtering results in a sample of atoms that are all in a single atomic state (state $|1\rangle$ in Fig. 1b) which allows adiabatic optical preparation of the atoms, as described below, and minimal heating of the cloud.

Next we evaporatively cool the atoms for 38 s to the transition temperature for Bose–Einstein condensation, T_c . The magnetic fields are then adjusted to adiabatically soften the trap. The resulting trapping potential has a frequency of $f_z = 21 \text{ Hz}$ along the symmetry (z) axis of the 4 Dee trap, and transverse frequencies $f_x = f_y = 69 \text{ Hz}$. The bias field, parallel to the z axis, is 11 G. When we cool well below T_c , we are left with 1–2 million atoms in the condensate. For these parameters the transition occurs at a temperature of $T_c = 435 \text{ nK}$ and a peak density in the cloud of $5 \times 10^{12} \text{ cm}^{-3}$.

We now apply a linearly polarized laser beam, the coupling beam, tuned to the transition between the unpopulated hyperfine states $|2\rangle$ and $|3\rangle$ (Fig. 1b). This beam couples states $|2\rangle$ and $|3\rangle$ and creates a quantum interference for a weaker probe laser beam (left circularly polarized) which is tuned to the $|1\rangle \rightarrow |3\rangle$ transition. A stable eigenstate (the ‘dark state’) of the atom in the presence of coupling and probe lasers is a coherent superposition of the two hyperfine ground states $|1\rangle$ and $|2\rangle$. The ratio of the probability amplitudes is such that the contributions to the atomic dipole moment induced by the two lasers exactly cancel. The quantum interference occurs in a narrow interval of probe frequencies, with a width determined by the coupling laser power.

Figure 2a shows the calculated transmission of the probe beam as a function of its detuning from resonance for parameters which are typical of this work. In the absence of dephasing of the $|1\rangle \rightarrow |2\rangle$ transition, the quantum interference would be perfect, and at line centre, the transmission would be unity. Figure 2b shows the refractive index for the probe beam as a function of detuning. Due to the very small Doppler broadening of the $|1\rangle \rightarrow |2\rangle$ transition in our nanokelvin samples, application of very low coupling intensity leads to a transparency peak with a width much smaller than the natural line width of the $|1\rangle \rightarrow |3\rangle$ transition. Correspondingly, the dispersion curve is much steeper than can be obtained by any other technique, and this results in the unprecedented low group velocities reported here. The group velocity v_g for a propagating electromagnetic pulse is^{16–19}.

$$v_g = \frac{c}{n(\omega_p) + \omega_p \frac{dn}{d\omega_p}} \approx \frac{\hbar c \epsilon_0 |\Omega_c|^2}{2\omega_p |\mu_{13}|^2 N} \quad (1)$$

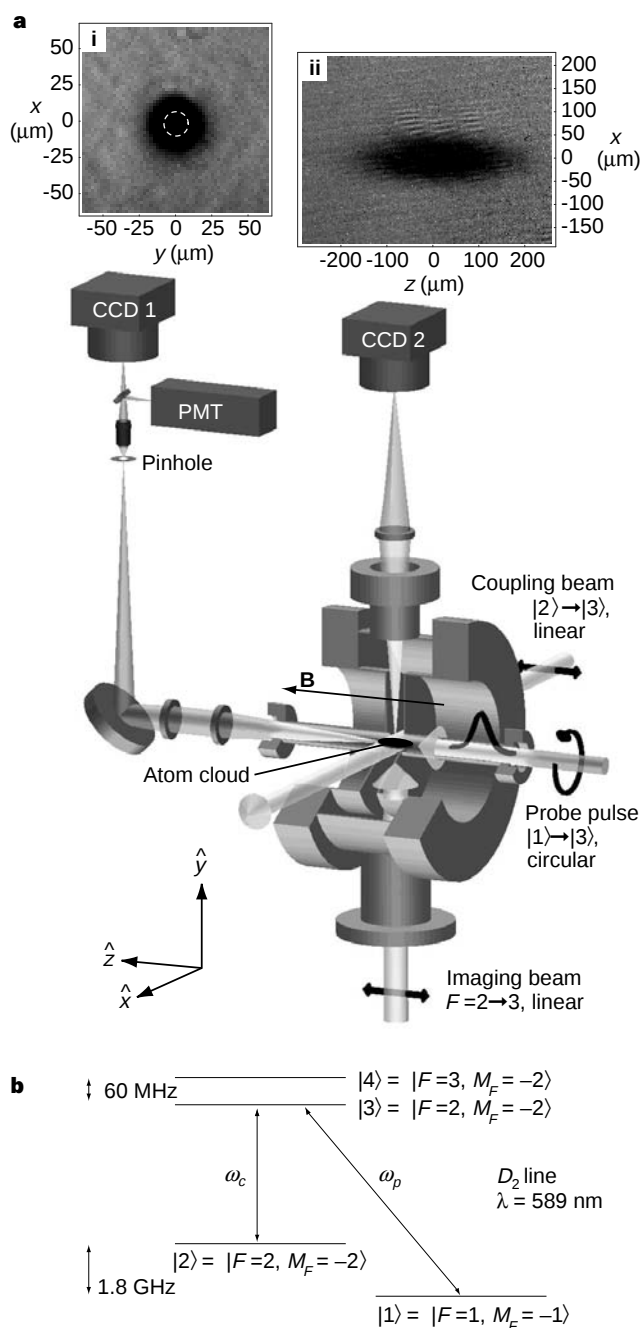


Figure 1 Experimental set-up. A ‘coupling’ laser beam propagates along the x axis with its linear polarization along the 11-G bias field in the z direction. The ‘probe’ laser pulse propagates along the z axis and is left-circularly polarized. With a flipper mirror in front of the camera CCD 1, we direct this probe beam either to the camera or to the photomultiplier (PMT). For pulse delay measurements, we place a pinhole in an external image plane of the imaging optics and select a small area, $15 \mu\text{m}$ in diameter, of the probe beam centred on the atom clouds (as indicated by the dashed circle in inset (i)). The pulse delays are measured with the PMT. The imaging beam propagating along the y axis is used to image atom clouds onto camera CCD 2 to find the length of the clouds along the pulse propagation direction (z axis) for determination of light speeds. Inset (ii) shows atoms cooled to 450 nK which is 15 nK above T_c . (Note that this imaging beam is never applied at the same time as the probe pulse and coupling laser). The position of a cloud and its diameter in the two transverse directions, x and y , are found with CCD 1. Inset (i) shows an image of a condensate.

Here $n(\omega_p)$ is the refractive index at probe frequency ω_p (rad s^{-1}), $|\Omega_c|^2$ is the square of the Rabi frequency for the coupling laser and varies linearly with intensity, μ_{13} is the electric dipole matrix element between states $|1\rangle$ and $|3\rangle$, N is the atomic density, and ϵ_0 is the permittivity of free space. At line centre, the refractive index is unity, and the second term in the denominator of equation (1) dominates the first. An important characteristic of the refractive index profile is that on resonance the dispersion of the group velocity is zero (see ref. 16), that is, $d^2n/d\omega_p^2 = 0$, and to lowest order, the pulse maintains its shape as it propagates. The established quantum interference allows pulse transmission through our atom clouds which would otherwise have transmission coefficients of e^{-10} (below T_c), and creates a steep dispersive profile and very low group velocity for light pulses propagating through the clouds.

We note that the centres of the curves in Fig. 2 are shifted by 0.6 MHz from probe resonance. This is due to a coupling of state $|2\rangle$ to state $|4\rangle$ through the coupling laser field, which results in an a.c. Stark shift of level $|2\rangle$ and a corresponding line shift of the $2 \rightarrow 3$ transition. As the transparency peak and unity refractive index are obtained at two-photon resonance, this leads to a refractive index at the $1 \rightarrow 3$ resonance frequency which is different from unity. The difference is proportional to the a.c. Stark shift and hence to the coupling laser intensity, which is important for predicting the nonlinear refractive index as described below.

A diagram of the experiment is shown in Fig. 1a. The 2.5-mm-diameter coupling beam propagates along the x axis with its linear polarization parallel to the $\mathbf{B}$ field. The 0.5-mm-diameter, σ^- polarized probe beam propagates along the z axis. The size and position of the atom cloud in the transverse directions, x and y , are obtained by imaging the transmission profile of the probe beam after the cloud onto a charge-coupled-device (CCD) camera. An image of a condensate is shown as inset (i). A 55 mW cm^{-2} coupling

laser beam was present during the 10- μs exposure of the atoms to a 5 mW cm^{-2} probe beam tuned close to resonance. The $f/7$ imaging optics are diffraction-limited to a resolution of $7 \mu\text{m}$.

During the pulse delay experiments, a pinhole (placed in an external image plane of the lens system) is used to select only the part of the probe light that has passed through the central $15 \mu\text{m}$ of the atom cloud where the column density is the greatest. The outline of the pinhole is indicated with the dashed circle in inset (i).

Both coupling and probe beams are derived from the same dye laser. The frequency of the coupling beam is set by an acousto-optic modulator (AOM) to the $|2\rangle \rightarrow |3\rangle$ resonance. Here we take into account both Zeeman shifts and the a.c. Stark shift described above.

The corresponding probe resonance is found by measuring the transmission of the probe beam as a function of its frequency. We apply a fast frequency sweep, across 32 MHz in 50 μs , and determine resonance from the transmission peak. The sweep is controlled by a separate AOM. The frequency is then fixed at resonance, and the temporal shape of the probe pulse is generated by controlling the r.f. drive power to the AOM. The resulting pulse is approximately gaussian with a full-width at half-maximum of 2.5 μs . The peak power is 1 mW cm^{-2} corresponding to a Rabi frequency of $\Omega_p = 0.20 A$, where the Einstein A coefficient is $6.3 \times 10^7 \text{ rad s}^{-1}$. To avoid distortion of the pulse, it is made of sufficient duration that its Fourier components are contained within the transparency peak.

Probe pulses are launched along the z axis 4 μs after the coupling beam is turned on (the coupling field is left on for 100 μs). Due to the magnetic filtering discussed above, all atoms are initially in state $|1\rangle$ which is a dark state in the presence of the coupling laser only. When the pulse arrives, the atoms adiabatically evolve so that the probability amplitude of state $|2\rangle$ is equal to the ratio $\Omega_p/(\Omega_p^2 + \Omega_c^2)^{1/2}$, where Ω_p is the probe Rabi frequency. To establish the coherent superposition state, energy is transferred from the

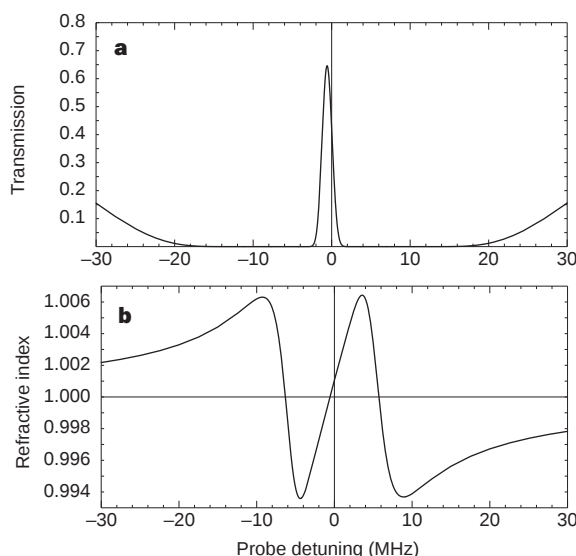


Figure 2 Effect of probe detuning. **a**, Transmission profile. Calculated probe transmission as a function of detuning from the $|1\rangle \rightarrow |3\rangle$ resonance for an atom cloud cooled to 450 nK, with a peak density of $3.3 \times 10^{12} \text{ cm}^{-3}$ and a length of $229 \mu\text{m}$ (corresponding to the cloud in inset (ii) of Fig. 1a). The coupling laser is resonant with the $|2\rangle \rightarrow |3\rangle$ transition and has a power density of 52 mW cm^{-2} . **b**, Refractive index profile. The calculated refractive index is shown as a function of probe detuning for the same parameters as in **a**. The steepness of the slope at resonance is inversely proportional to the group velocity of transmitted light pulses and is controlled by the coupling laser intensity. Note that as a result of the a.c. Stark shift of the $|2\rangle \rightarrow |3\rangle$ transition, caused by a coupling of states $|2\rangle$ and $|4\rangle$ through the coupling laser field, the centre of the transmission and refractive index profiles is shifted by 0.6 MHz. The shift of the refractive index profile results in the nonlinear refractive index described in the text.

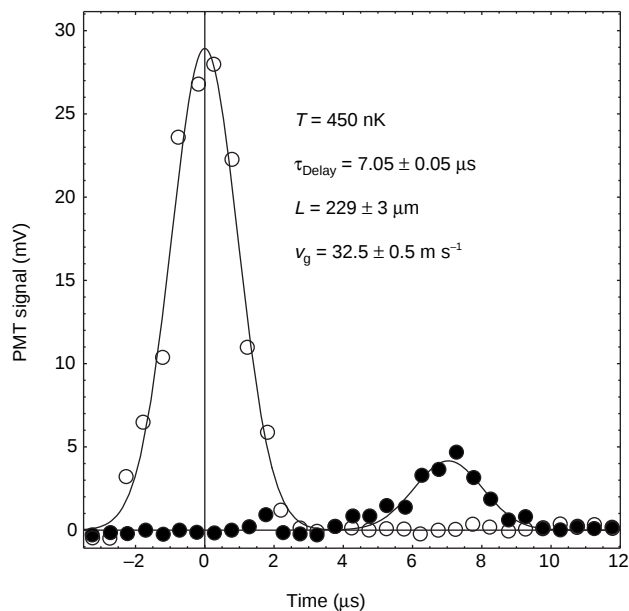


Figure 3 Pulse delay measurement. The front pulse (open circles) is a reference pulse with no atoms in the system. The other pulse (filled circles) is delayed by $7.05 \mu\text{s}$ in a $229 \mu\text{m}$ -long atom cloud (see inset (ii) in Fig. 1a). The corresponding light speed is 32.5 m s^{-1} . The curves represent gaussian fits to the measured pulses.

front of the probe pulse to the atoms and the coupling laser field. At the end of the pulse, the atoms adiabatically return to the original state $|1\rangle$ and the energy returns to the back of the probe pulse with no net energy and momentum transfer to the atomic cloud. Because the refractive index is unity, the electric field is unchanged as the probe pulse enters the medium. As the group velocity is decreased, the total energy density must increase so as to keep constant the power per area. This increase is represented by the energy stored in the atoms and the coupling laser field during pulse propagation through the cloud.

The pulses are recorded with a photomultiplier (3-ns response time) after they penetrate the atom clouds. The output from the photomultiplier is amplified by a 150-MHz-bandwidth amplifier and the waveforms are recorded on a digital scope. With a 'flipper' mirror in front of the camera we control whether the probe beam is directed to the camera or to the photomultiplier.

The result of a pulse delay measurement is shown in Fig. 3. The front pulse is a reference pulse obtained with no atoms present. The pulse delayed by $7.05 \mu\text{s}$ was slowed down in an atom cloud with a length of $229 \mu\text{m}$ (see Fig. 1a, inset (ii)). The resulting light speed is 32.5 m s^{-1} . We used a coupling laser intensity of 12 mW cm^{-2} corresponding to a Rabi frequency of $\Omega_c = 0.56 A$. The cloud was cooled to 450 nK (which is 15 nK above T_c), the peak density was $3.3 \times 10^{12} \text{ cm}^{-3}$, and the total number of atoms was 3.8×10^6 . From these numbers we calculate that the pulse transmission coefficient would be e^{-63} in the absence of the coupling laser. The probe pulse was indeed observed to be totally absorbed by the atoms when the coupling beam was left off. Inhomogeneous broadening due to spatially varying Zeeman shifts is negligible ($\sim 20 \text{ kHz}$) for the low temperatures and correspondingly small cloud sizes used here.

The size of the atom cloud in the z direction is obtained with another CCD camera. For this purpose, we use a separate 1 mW cm^{-2} laser beam propagating along the vertical y axis and tuned 20 MHz below the $F = 2 \rightarrow 3$ transition. The atoms are pumped to the

$F = 2$ ground state for $10 \mu\text{s}$ before the imaging which is performed with an exposure time of $10 \mu\text{s}$. We image the transmission profile of the laser beam after the atom cloud with diffraction-limited $f/5$ optics. An example is shown in Fig. 1a, inset (ii), where the asymmetry of the trap is clear from the cloud's elliptical profile. We note that the imaging laser is never applied at the same time as the coupling laser and probe pulse, and for each recorded pulse or CCD picture a new cloud is loaded.

We measured a series of pulse delays and corresponding cloud sizes for atoms cooled to temperatures between $2.5 \mu\text{K}$ and 50 nK . From these pairs of numbers we obtain the corresponding propagation velocities (Fig. 4). The open circles are for a coupling power of 52 mW cm^{-2} ($\Omega_c = 1.2 A$). The light speed is inversely proportional to the atom density (equation (1)) which increases with lower temperatures, with an additional density increase when a condensate is formed. The filled circles are for a coupling power of 12 mW cm^{-2} . The lower coupling power is seen to cause a decrease of group velocities in agreement with equation (1). We obtain a light speed of 17 m s^{-1} for pulse propagation in an atom cloud initially prepared as an almost pure Bose-Einstein condensate (condensate fraction is $\geq 90\%$). Whether the cloud remains a condensate during and after pulse propagation is an issue that is beyond the scope of this Letter.

Transitions from state $|2\rangle$ to state $|4\rangle$, induced by the coupling laser (detuned by 60 MHz from this transition), result in a finite decay rate of the established coherence between states $|1\rangle$ and $|2\rangle$ and limit pulse transmission. The dephasing rate is proportional to the power density of the coupling laser and we expect, and find, that probe pulses have a peak transmission that is independent of coupling intensity and a velocity which reduces linearly with this intensity. The dephasing time is determined from the slope of a semi-log plot of transmission versus pulse delay¹⁹. At a coupling power of 12 mW cm^{-2} , we measured a dephasing time of $9 \mu\text{s}$ for atom clouds just above T_c .

Giant Kerr nonlinearities are of interest for areas of quantum optics such as optical squeezing, quantum nondemolition, and studies of nonlocality. It was recently proposed that they may be obtained using electromagnetically induced transparency²⁰. Here we report the first (to our knowledge) measurement of such a nonlinearity. The refractive index for zero probe detuning is given by $n = 1 + (n_2 I_c)$ where I_c is the coupling laser intensity, and n_2 the cross phase nonlinear refractive index. As seen from Fig. 2b, the nonlinear term ($n_2 I_c$) equals the product of the slope of the refractive index at probe resonance and the a.c. Stark shift of the $|2\rangle \rightarrow |3\rangle$ transition caused by the coupling laser. We can then express n_2 by the formula (see equation (1));

$$n_2 = \frac{\Delta\omega_s}{I_c} \frac{dn}{d\omega_p} \approx \frac{1}{2\pi} \frac{\Delta\omega_s}{I_c} \frac{\lambda}{v_g} \quad (2)$$

where $\Delta\omega_s$ is the a.c. Stark shift, proportional to I_c , and λ the wavelength of the probe transition. We measured an a.c. Stark shift of $1.3 \times 10^6 \text{ rad s}^{-1}$ for a coupling laser intensity of 40 mW cm^{-2} . For a measured group velocity of 17 m s^{-1} (Fig. 4), we obtain a nonlinear refractive index of $0.18 \text{ cm}^2 \text{ W}^{-1}$. This nonlinear index is $\sim 10^6$ times greater than that measured in cold Cs atoms²¹.

With a system that avoids the $|1\rangle$ - $|2\rangle$ dephasing rate described above (which can be obtained by tuning to the D_1 line in sodium), the method used here could be developed to yield the collision-induced dephasing rate of the double condensate which is generated in the process of establishing electromagnetically induced transparency (see also refs 22, 23). In that case, the square of the probability amplitude for state $|3\rangle$ could be kept below 10^{-5} during pulse propagation, with no heating of the condensate as a result. With improved frequency stability of our set-up and lower coupling intensities, even lower light speeds would be possible, perhaps of the order of centimetres per second, comparable to the speed of

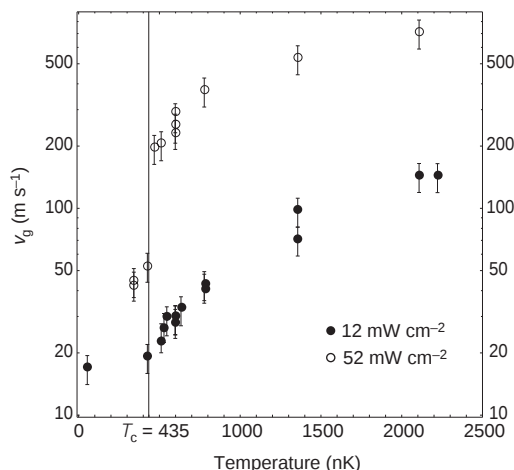


Figure 4 Light speed versus atom cloud temperature. The speed decreases with temperature due to the atom density increase. The open circles are for a coupling power of 52 mW cm^{-2} and the filled circles are for a coupling power of 12 mW cm^{-2} . The temperature T_c marks the transition temperature for Bose-Einstein condensation. The decrease in group velocity below T_c is due to a density increase of the atom cloud when the condensate is formed. From imaging measurements we obtain a maximum atom density of $8 \times 10^{13} \text{ cm}^{-3}$ at a temperature of 200 nK . Here, the dense condensate component constitutes 60% of all atoms, and the total atom density is 16 times larger than the density of a non-condensed cloud at T_c . The light speed measurement at 50 nK is for a cloud with a condensate fraction $\geq 90\%$. The finite dephasing rate due to state $|4\rangle$ does not allow pulse penetration of the most dense clouds. This problem could be overcome by tuning the laser to the D_1 line as described in the text.

sound in a Bose–Einstein condensate. Under these conditions we expect phonon excitation during light pulse propagation through the condensate. By deliberately tuning another laser beam to the $|2\rangle \rightarrow |4\rangle$ transition, it should be possible to demonstrate optical switching at the single photon level²⁴. Finally, we note that during propagation of the atom clouds, light pulses are compressed in the z direction by a ratio of c/v_g . For our experimental parameters, that results in pulses with a spatial extent of only 43 μm . $\square$

Received 3 November; accepted 21 December 1998.

1. Knight, P. L., Stoicheff, B. & Walls, D. (eds) Highlights in quantum optics. *Phil. Trans. R. Soc. Lond. A* **355**, 2215–2416 (1997).
2. Harris, S. E. Electromagnetically induced transparency. *Phys. Today* **50**(7), 36–42 (1997).
3. Scully, M. O. & Zubairy, M. S. *Quantum Optics* (Cambridge Univ. Press, 1997).
4. Arimondo, E. in *Progress in Optics* (ed. Wolf, E.) 257–354 (Elsevier Science, Amsterdam, 1996).
5. Bergmann, K., Theuer, H. & Shore, B. W. Coherent population transfer among quantum states of atoms and molecules. *Rev. Mod. Phys.* **70**, 1003–1006 (1998).
6. Chu, S. The manipulation of neutral particles. *Rev. Mod. Phys.* **70**, 685–706 (1998).
7. Cohen-Tannoudji, C. N. Manipulating atoms with photons. *Rev. Mod. Phys.* **70**, 707–719 (1998).
8. Phillips, W. D. Laser cooling and trapping of neutral atoms. *Rev. Mod. Phys.* **70**, 721–741 (1998).
9. Hess, H. F. Evaporative cooling of magnetically trapped and compressed spin-polarized hydrogen. *Phys. Rev. B* **34**, 3476–3479 (1986).
10. Masuhara, N. *et al.* Evaporative cooling of spin-polarized atomic hydrogen. *Phys. Rev. Lett.* **61**, 935–938 (1988).
11. Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Observation of Bose-Einstein condensation in a dilute atomic vapor. *Science* **269**, 198–201 (1995).
12. Davis, K. B. *et al.* Bose-Einstein condensation in a gas of sodium atoms. *Phys. Rev. Lett.* **75**, 3969–3973 (1995).
13. Bradley, C. C., Sackett, C. A. & Hulet, R. G. Bose-Einstein condensation of lithium: observation of limited condensate number. *Phys. Rev. Lett.* **78**, 985–989 (1997).
14. Hau, L. V. *et al.* Near-resonant spatial images of confined Bose-Einstein condensates in a 4-Dee magnetic bottle. *Phys. Rev. A* **58**, R54–R57 (1998).
15. Hau, L. V., Golovchenko, J. A. & Burns, M. M. A new atomic beam source: The “candlestick”. *Rev. Sci. Instrum.* **65**, 3746–3750 (1994).
16. Harris, S. E., Field, J. E. & Kasapi, A. Dispersive properties of electromagnetically induced transparency. *Phys. Rev. A* **46**, R29–R32 (1992).
17. Grobe, R., Hioe, F. T. & Eberly, J. H. Formation of shape-preserving pulses in a nonlinear adiabatically integrable system. *Phys. Rev. Lett.* **73**, 3183–3186 (1994).
18. Xiao, M., Li, Y.-Q., Jin, S.-Z. & Gea-Banacloche, J. Measurement of dispersive properties of electromagnetically induced transparency in rubidium atoms. *Phys. Rev. Lett.* **74**, 666–669 (1995).
19. Kasapi, A., Jain, M., Yin, G. Y. & Harris, S. E. Electromagnetically induced transparency: propagation dynamics. *Phys. Rev. Lett.* **74**, 2447–2450 (1995).
20. Schmidt, H. & Imamoglu, A. Giant Kerr nonlinearities obtained by electromagnetically induced transparency. *Opt. Lett.* **21**, 1936–1938 (1996).
21. Lambrecht, A., Courty, J. M., Reynaud, S. & Giacobino, E. Cold atoms: A new medium for quantum optics. *Appl. Phys. B* **60**, 129–134 (1995).
22. Hall, D. S., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Measurements of relative phase in two-component Bose-Einstein condensates. *Phys. Rev. Lett.* **81**, 1543–1546 (1998).
23. Ruostekoski, J. & Walls, D. F. Coherent population trapping of Bose-Einstein condensates: detection of phase diffusion. *Eur. Phys. J. D* (submitted).
24. Harris, S. E. & Yamamoto, Y. Photon switching by quantum interference. *Phys. Rev. Lett.* **81**, 3611–3614 (1998).

Acknowledgements. We thank J. A. Golovchenko for discussions and C. Liu for experimental assistance. L.V.H. acknowledges support from the Rowland Institute for Science. S.E.H. is supported by the US Air Force Office of Scientific Research, the US Army Research Office, and the US Office of Naval Research. C.H.B. is supported by an NSF fellowship.

Correspondence and requests for materials should be addressed to L.V.H. (e-mail: hau@rowland.org).

Luttinger-liquid behaviour in carbon nanotubes

Marc Bockrath*, David H. Cobden*, Jia Lu*, Andrew G. Rinzier†, Richard E. Smalley†, Leon Balents‡ & Paul L. McEuen*

* Department of Physics, University of California and Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

† Center for Nanoscale Science and Technology, Rice Quantum Institute and Department of Chemistry and Physics, MS-100, Rice University, PO Box 1892, Houston, Texas 77251, USA

‡ Institute for Theoretical Physics, University of California, Santa Barbara, California 93106-4030, USA

Electron transport in conductors is usually well described by Fermi-liquid theory, which assumes that the energy states of the electrons near the Fermi level E_F are not qualitatively altered by Coulomb interactions. In one-dimensional systems, however, even weak Coulomb interactions cause strong perturbations.

The resulting system, known as a Luttinger liquid, is predicted to be distinctly different from its two- and three-dimensional counterparts¹. For example, tunnelling into a Luttinger liquid at energies near the Fermi level is predicted to be strongly suppressed, unlike in two- and three-dimensional metals. Experiments on one-dimensional semiconductor wires^{2,3} have been interpreted by using Luttinger-liquid theory, but an unequivocal verification of the theoretical predictions has not yet been obtained. Similarly, the edge excitations seen in fractional quantum Hall conductors are consistent with Luttinger-liquid behaviour^{4,5}, but recent experiments failed to confirm the predicted relationship between the electrical properties of the bulk state and those of the edge states⁶. Electrically conducting single-walled carbon nanotubes (SWNTs) represent quantum wires^{7–10} that may exhibit Luttinger-liquid behaviour^{11,12}. Here we present measurements of the conductance of bundles (‘ropes’) of SWNTs as a function of temperature and voltage that agree with predictions for tunnelling into a Luttinger liquid. In particular, we find that the conductance and differential conductance scale as power laws with respect to temperature and bias voltage, respectively, and that the functional forms and the exponents are in good agreement with theoretical predictions.

SWNTs are sufficiently robust and long to allow electrical connections to lithographically defined metallic electrodes, thereby making it possible to probe the intriguing electrical properties of

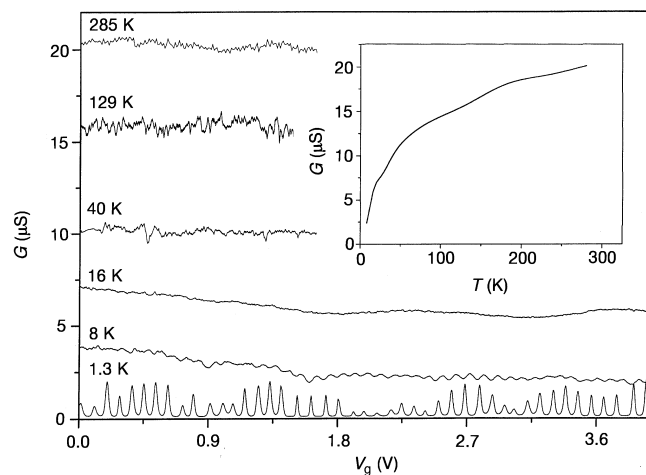


Figure 1 The two-terminal linear-response conductance G versus gate voltage V_g for a bulk-contacted metallic nanotube rope at a variety of temperatures. The data show significant temperature dependence for energy scales above the charging energy that cannot be explained by the Coulomb blockade model. Inset: average conductance as a function of temperature T . The samples used in these experiments are made in one of two ways. In both methods, SWNTs are deposited from a suspension in dichloroethane onto a 1- μm -thick layer of SiO_2 that has been thermally grown on a degenerately doped Si wafer, used as a gate electrode. Atomic force microscopy imaging reveals that the diameters of the ropes vary between 1 and 10 nm. In the first method⁹, chromium-gold contacts are applied over the top of the nanotube rope using electron beam lithography and lift-off. From measurements of these devices in the Coulomb blockade regime, we conclude that the electrons are confined to the length of rope between the leads. This implies that the leads cut the nanotubes into segments, and transport involves tunnelling into the ends of the nanotubes (‘end-contacted’). In the second method¹⁰, electron-beam lithography is first used to define leads, and ropes are deposited on top of the leads. Samples were selected that showed Coulomb blockade behaviour at low temperatures with a single well-defined period, indicating the presence of a single quantum dot. The charging energy of these samples indicates a quantum dot with a size substantially larger than the spacing between the leads, as found by Tans *et al.*¹⁰. Transport thus occurs by electrons tunnelling into the middle, or bulk, of the nanotubes (‘bulk-contacted’).

these nanometre-sized structures^{7–10,13,14}. Individual tubes, for example, are either semiconducting or conducting, depending on their chirality^{7,8}, whereas electron transport through ropes is typically dominated by a single metallic nanotube within the rope⁹. The latter observation is in agreement with the finding that most of the nanotubes in a rope are semiconductors and thus insulating at the low temperatures of transport measurements^{7,8}.

Electrical connections to nanotubes and nanotube ropes can be achieved by either depositing electrode metal over the top of the tubes ('end-contacted' samples), or by placing the tubes on top of predefined metal leads ('bulk-contacted' samples). We use both geometries to study the transport properties of nanotube ropes. Figure 1 gives an example of the measured two-terminal conductance, G , as a function of gate voltage, V_g , for a bulk-contacted metallic rope at different temperatures. The Coulomb oscillations¹⁵ that occur each time an electron is added to a nanotube within the rope are clearly visible at low temperatures. The temperature dependence of the oscillations yields a charging energy U for this sample of 1.9 meV. At temperatures above 20 K, the thermal energy exceeds the charging energy (that is, $k_B T > U$, where k_B is Boltzmann's constant and T the absolute temperature). This results in the Coulomb oscillations being nearly completely 'washed out', rendering the conductance independent of gate voltage. The dependence on temperature is illustrated in Fig. 1 inset, which shows the conductance dropping steeply as the temperature is lowered, extrapolating to $G = 0$ at $T = 0$.

Figure 2 shows G as a function of T on a double logarithmic scale for two bulk-contacted and two end-contacted nanotube ropes (Fig. 2a and b, respectively). The measured data (solid lines) show approximate power-law behaviour, $G \propto T^\alpha$, for the four samples shown. However, the range of temperature over which this behaviour occurs is limited by the effects of Coulomb blockade at low

temperatures. After correcting for the known temperature dependence due to the Coulomb blockade¹⁵, the corrected data (dashed lines) show power-law behaviour over a greater temperature range, with slightly different exponents. Above $T \approx 100$ K, G begins to saturate for some samples. This saturation is observed in many, but not all, of the samples studied.

The corrected data obtained for the bulk-contacted samples show approximate power-law behaviour from 8 to 300 K with exponents $\alpha_{\text{bulk}} \approx 0.33$ and 0.38. In the case of the end-contacted samples, the corrected data show approximate power-law behaviour from 10 to 100 K with exponents $\alpha_{\text{end}} \approx 0.6$ for both samples. The upper inset to Fig. 2a shows the exponents determined from the temperature dependence of a variety of samples. Exponents marked with 'x' and 'o' are for bulk- and end-contacted tubes, respectively. The former show a systematically lower exponent than the latter samples with $\alpha_{\text{end}} \approx 0.6$ and $\alpha_{\text{bulk}} \approx 0.3$.

The two insets in Fig. 3 show the measured differential conductance dI/dV as a function of the applied bias voltage V . The upper inset in Fig. 3a shows results for a bulk-contacted sample (see lower inset) at different temperatures, plotted on a log-log scale. In linear response, dI/dV is proportional to a (temperature-dependent) constant, $G(T)$ from Fig. 2. At high biases, dI/dV increases with increasing V . The curves at different temperatures fall onto a single curve in the high-bias regime. As this curve is roughly linear on a log-log plot, it implies that the differential conductance is described by a power law, $dI/dV \propto V^\alpha$, where $\alpha = 0.36$. At the lowest temperature $T = 1.6$ K, this power-law behaviour occurs over two decades in V , from 1 to 100 mV.

The upper inset in Fig. 3b shows dI/dV as a function of V for an end-contacted sample (see lower inset) at several temperatures. The conductance is again a temperature-dependent constant at low biases $eV \ll k_B T$, whereas at higher biases dI/dV increases. The

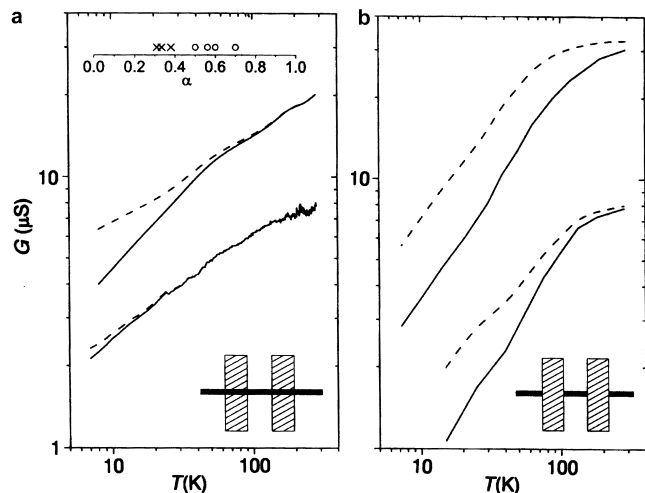


Figure 2 Conductance G plotted against temperature T for individual nanotube ropes. The data are plotted on a log-log scale. **a**, Data for ropes that are deposited over pre-defined leads (bulk-contacted); **b**, data for ropes that are contacted by evaporating the leads on top of the ropes (end-contacted). Sketches depicting the measurement configuration are shown in the lower insets. The plots show both the raw data (solid line) and the data corrected for the temperature dependence expected from the Coulomb blockade (CB) model (dashed line). We correct the data by dividing the measured $G(T)$ by the theoretically expected temperature dependence in the CB model. This correction factor depends only on $U/k_B T$, and, because U can be independently measured from the temperature dependence of the Coulomb oscillations, the correction procedure requires no adjustable parameters. If the CB were the only source of the temperature dependence, the dashed lines would be horizontal. Instead they have a finite slope, indicating an approximate power-law dependence on T . The upper inset to **a** shows the power-law exponents inferred for a variety of samples. Open circles denote end-contacted samples, and crosses denote bulk-contacted ones.

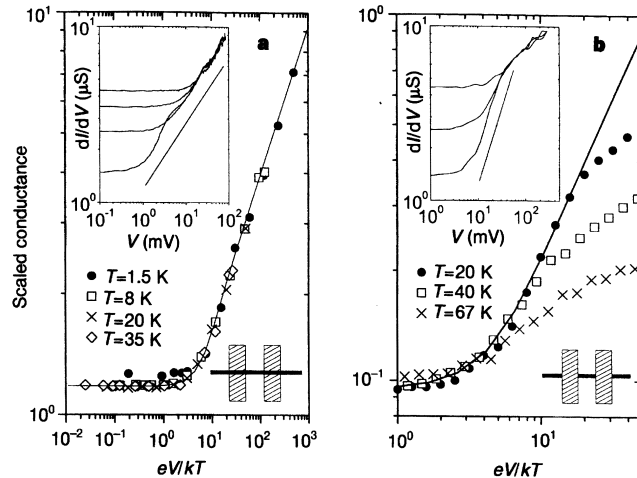


Figure 3 The differential conductance dI/dV measured at various temperatures. Inset in **a**, dI/dV curves taken on a bulk-contacted rope at temperatures $T = 1.6$ K, 8 K, 20 K and 35 K. Inset in **b**, dI/dV curves taken on an end-contacted rope at temperatures $T = 20$ K, 40 K and 67 K. In both insets, a straight line on the log-log plot is shown as a guide to the eye to indicate power-law behaviour. The main panels **a** and **b** show these measurements collapsed onto a single curve by using the scaling relations described in the text. The solid line is the theoretical result fitted to the data by using γ as a fitting parameter. The values of γ resulting in the best fit to the data are $\gamma = 0.46$ in **a** and $\gamma = 0.63$ in **b**.

high-bias data follows an approximate power law before rolling off to a reduced slope for $V > 30$ mV. Although the range of data is too small to conclude that a power law accurately describes the behaviour at intermediate voltages, if a straight line is fitted to the range $9 \text{ mV} < V < 32 \text{ mV}$, the exponent obtained is $\alpha = 0.87$.

A possible explanation for the approximate power-law behaviour that we see in our data is a strong energy dependence of the tunnel barrier, with increased tunnelling efficiency at high energies. This would lead to activated transport over the barrier, so that the conductance can be described by $G \propto \exp(-V_b/k_B T)$, where V_b is the height of the tunnel barriers. However, this general expression for G is inconsistent with our finding that the conductance extrapolates to $G = 0$ at $T = 0$ (Fig. 1 inset). The type of power-law behaviour that we observe could also arise if the electron transport were to occur through multiple quantum dots in series. Multiple quantum dots can be formed by disorder¹⁶ or by barriers produced when the nanotubes bend over the lithographically defined contacts¹⁷. But as we have chosen to study only nanotube ropes that exhibit a single dominant period for the Coulomb oscillations at low temperatures, our samples are likely to contain only a single quantum dot.

A third possible explanation for the experimental observations would be that the nanotube ropes behave as a Luttinger liquid (LL). An LL is a one-dimensional correlated electron state characterized by a parameter g that measures the strength of the interaction between electrons. Strong repulsive interactions are characterized by $g \ll 1$, whereas $g = 1$ for the non-interacting electron gas. However, for any $g \neq 1$, the low-energy excitations of the system are not all weakly interacting quasiparticles, and the Fermi liquid theory used to describe conventional metals is not appropriate.

In SWNTs, the long-range Coulomb interaction between electrons is expected to yield an LL with $g < 1$ (refs 11, 12). For a finite-length tube or rope, the Luttinger parameter g is given by:

$$g = \left[1 + \frac{2U}{\Delta} \right]^{-1/2} \quad (1)$$

where U is the charging energy of the tube and Δ is the single-particle level spacing (the two one-dimensional sub-bands of the nanotube are assumed to be non-degenerate). From previous measurements and theoretical estimates^{9,10} $U/\Delta \approx 6$, yielding an expected Luttinger parameter $g_{\text{theory}} \approx 0.28$.

The tunnelling of an electron into an LL is dramatically different from tunnelling into Fermi liquid. For a Fermi liquid, an energy-independent tunnelling amplitude is expected. This yields a temperature- and bias-independent tunnelling conductance. For a clean LL, on the other hand, the tunnelling amplitude is predicted to vanish as a power law in the energy of the tunnelling electron. This leads to a power-law variation of G with T at small biases ($eV \ll k_B T$);

$$G(T) \propto T^\alpha \quad (2)$$

or, with V at large biases ($eV \gg k_B T$):

$$dI/dV \propto V^\alpha \quad (3)$$

The exponent of these power laws depends on the number of one-dimensional channels¹⁸ and on whether the electron tunnels into the bulk or the end of the LL. For a SWNT with four conducting modes at E_F , the exponents are^{11,12}:

$$\alpha_{\text{end}} = (g^{-1} - 1)/4 \quad (4a)$$

$$\alpha_{\text{bulk}} = (g^{-1} + g - 2)/8 \quad (4b)$$

Using equations (1) and (4), we obtain $\alpha_{\text{end}}(\text{theory}) = 0.65$ and $\alpha_{\text{bulk}}(\text{theory}) = 0.24$.

To compare these theoretical predictions for tunnelling into an isolated nanotube through a single barrier to the experimental data obtained for ropes connected by two contacts, we must make two

assumptions. First, we assume that transport in the rope is dominated by a single metallic tube, as discussed previously. Preliminary theoretical studies (L.B. and C. Kane, manuscript in preparation) of ropes composed of SWNTs with a relatively small fraction of metallic tubes support this assumption. These studies find that the only significant inter-tube coupling is electrostatic. Such an interaction will introduce extra screening of the Coulomb interaction but, because of the weak (logarithmic) dependence of g on the screening length, the LL predictions are essentially unchanged. Second, we assume that the tunnel resistances into and out of the tube are the dominant resistances in the system. The circuit thus consists of two tunnel junctions in series, with the current response of each junction described by equations (1)–(4). We note that the voltage drop across the highest-impedance junction will be some fraction γ of the total applied bias V , where $1/2 \leq \gamma \leq 1$. If the barriers are equal, the voltage will divide equally between these junctions and $\gamma = 1/2$. Alternatively, if the resistance of one junction dominates, $\gamma = 1$.

With these assumptions, the approximate power-law behaviour as a function of T or V observed in Figs 2 and 3 then follows from equations (1)–(4). The predicted values of the exponents are in good agreement with the experimental values. This agreement may be somewhat fortuitous owing to the experimental uncertainty in the value of U/Δ and complexities associated with the screening of the Coulomb interaction by the metallic leads^{11,12}. Nevertheless, the measurements are described both qualitatively and quantitatively by LL theory. Power-law behaviour in T is observed up to 300 K in the bulk-contacted samples, indicating that nanotubes are LLs even at room temperature.

At present, we do not understand the origins of the high-energy saturation observed in the end-contacted tubes. One possibility is that, at high energies, electrons can tunnel in both directions and hence the end-contacted tubes behave as bulk-contacted tubes, with a correspondingly lower exponent. Further experiments are necessary to clarify this issue.

The LL theory makes an additional prediction for this system. The differential conductance for a single tunnel junction is given by a universal scaling curve^{19,20}:

$$\frac{dI}{dV} = AT^\alpha \cosh\left(\gamma \frac{eV}{2k_B T}\right) \left| \Gamma\left(\frac{1+\alpha}{2} + \gamma \frac{ieV}{2\pi k_B T}\right) \right|^2 \quad (5)$$

where $\Gamma(x)$ is the gamma function, γ is the constant introduced earlier that takes into account the voltage division between the two tunnel junctions, and A is an arbitrary constant. This equation assumes that the leads are at $T = 0$ K. For leads at a finite temperature, dI/dV is given by the convolution of equation (5) and the derivative of the Fermi distribution: $df/dE = (1/4k_B T) \text{sech}^2(\gamma eV/2k_B T)$.

If the above scaling relation is correct, it should be possible to collapse the data at different temperatures onto a single universal curve. To do this, the measured dI/dV at each temperature was divided by T^α and plotted against $eV/k_B T$, as shown in Figs 3a and b. For both geometries, the scaled conductance is constant as $eV/k_B T$ approaches zero, but increases when $eV/k_B T$ is significantly greater than one. The data collapse quite well onto a universal curve for the bulk-contacted device over the entire bias range (Fig. 3a). For the end-contacted device, the data deviate from power-law behaviour for biases $V > 30$ mV, as discussed previously. This is reflected in Fig. 3b in a roll-off that occurs at lower values of $eV/k_B T$ as the temperature is increased.

The solid lines in Fig. 3a and b are a plot of the curve obtained by fitting equation (5) (convolved with df/dE) to the data, with γ as a fitting parameter. The theory fits the scaled data reasonably well, especially for the bulk-contacted tube, and yields $\gamma = 0.5 \pm 0.1$ and $\gamma = 0.6 \pm 0.1$ for the bulk-contacted and end-contacted tubes, respectively. Given the associated experimental uncertainty, these values fall within the allowable range ($0.5 < \gamma < 1$) for two barriers in series.

Taken as a whole, the data shown in Figs 2 and 3 provide strong evidence that the electrons in metallic carbon nanotubes constitute an LL. Future work will test other predictions of this theory, such as tunnelling between LLs in end-to-end¹ and in crossed geometries²¹. □

Received 2 July; accepted 1 December 1998.

1. Fisher, M. P. A. & Glazman, A. *Mesoscopic Electron Transport* (Kluwer Academic, Boston, 1997).
2. Tarucha, S., Honda, T. & Saku, T. Reduction of quantized conductance at low temperatures observed in 2 to 10 μm -long quantum wires. *Solid State Commun.* **94**, 413–418 (1995).
3. Yacoby, A. *et al.* Nonuniversal conductance quantization in quantum wires. *Phys. Rev. Lett.* **77**, 4612–4615 (1996).
4. Milliken, F. P., Umbach, C. P. & Webb, R. A. Indications of a Luttinger liquid in the fractional quantum Hall regime. *Solid State Commun.* **97**, 309–313 (1996).
5. Chang, A. M., Pfeiffer, L. N. & West, K. W. Observation of chiral Luttinger behavior in electron tunneling into fractional quantum Hall edges. *Phys. Rev. Lett.* **77**, 2538–2541 (1996).
6. Grayson, M. *et al.* Continuum of chiral Luttinger liquids at the fractional quantum Hall edge. *Phys. Rev. Lett.* **80**, 1062–1065 (1998).
7. Wildoer, J. W. G. *et al.* Electronic structure of atomically resolved carbon nanotubes. *Nature* **391**, 59–62 (1998).
8. Odom, T. W., Huang, J., Kim, P. & Lieber, C. M. Atomic structure and electronic properties of single-walled carbon nanotubes. *Nature* **391**, 62–64 (1998).
9. Bockrath, M. *et al.* Single-electron transport in ropes of carbon nanotubes. *Science* **275**, 1922–1925 (1997).
10. Tans, S. J. *et al.* Individual single-wall carbon nanotubes as quantum wires. *Nature* **386**, 474–477 (1997).
11. Egger, R. & Gogolin, A. Effective low-energy theory for correlated carbon nanotubes. *Phys. Rev. Lett.* **79**, 5082–5085 (1997).
12. Kane, C., Balents, L. & Fisher, M. P. A. Coulomb interactions and mesoscopic effects in carbon nanotubes. *Phys. Rev. Lett.* **79**, 5086–5089 (1997).
13. Cobden, D. H. *et al.* Spin splitting and even–odd effects in carbon nanotubes. *Phys. Rev. Lett.* **81**, 681–684 (1998).
14. Tans, S. J., Devoret, M. H., Groeneveld, R. J. A. & Dekker, C. Electron–electron correlations in carbon nanotubes. *Nature* **394**, 761–764 (1998).
15. Grabert, H. & Devoret, M. H. *Single Charge Tunneling: Coulomb Blockade Phenomena in Nanostructures* (Plenum, New York, 1992).
16. Kane, C. L. & Mele, E. J. Size, shape, and low energy electronic structure of carbon nanotubes. *Phys. Rev. Lett.* **78**, 1932–1935 (1997).
17. Bezryadin, A., Verschuere, A. R. M., Tans, S. J. & Dekker, C. Multiprobe transport experiments on individual single-wall carbon nanotubes. *Phys. Rev. Lett.* **80**, 4036–4039 (1998).
18. Matveev, K. A. & Glazman, L. I. Coulomb blockade of tunneling into a quasi-one-dimensional wire. *Phys. Rev. Lett.* **70**, 990–993 (1993).
19. Fisher, M. P. A. & Dorsey, A. Dissipative quantum tunneling in a biased double-well system at finite temperatures. *Phys. Rev. Lett.* **54**, 1609–1612 (1985).
20. Grabert, H. & Weiss, U. Quantum tunneling rates for asymmetric double-well systems with ohmic dissipation. *Phys. Rev. Lett.* **54**, 1605–1608 (1985).
21. Komnik, A. & Egger, R. Nonequilibrium transport for crossed Luttinger liquids. *Phys. Rev. Lett.* **80**, 2881–2884 (1998).

Acknowledgements. We thank S. Louie, M. Cohen, D.-h. Lee, A. Zettl and A. Georges for discussions. This work was supported by DOE (Basic Energy Sciences, Materials Sciences Division, the sp^2 Materials Initiative). L.B. was supported by the NSF.

Correspondence and requests for materials should be addressed to P.L.M. (mceuen@socrates.berkeley.edu).

The nature of the hydrated excess proton in water

Dominik Marx*, Mark E. Tuckerman†, Jürg Hutter* & Michele Parrinello*

* Max-Planck-Institut für Festkörperforschung, Heisenbergstrasse 1, 70569 Stuttgart, Germany

† Department of Chemistry and Courant Institute of Mathematical Sciences, New York University, 4 Washington Place, New York, New York 10003, USA

Explanations for the anomalously high mobility of protons in liquid water began with Grotthuss's idea^{1,2} of 'structural diffusion' nearly two centuries ago. Subsequent explanations have refined this concept by invoking thermal hopping^{3,4}, proton tunnelling^{5,6} or solvation effects⁷. More recently, two main structural models have emerged for the hydrated proton. Eigen^{8,9} proposed the formation of an H_3O_4^+ complex in which an H_3O^+ core is strongly hydrogen-bonded to three H_2O molecules. Zundel^{10,11}, meanwhile, supported the notion of an H_5O_2^+ complex in which the proton is shared between two H_2O molecules. Here we use *ab initio* path integral^{12–14} simulations to address this question. These simulations include time-independent equilibrium thermal and quantum fluctuations of all nuclei, and determine interatomic

interactions from the electronic structure. We find that the hydrated proton forms a fluxional defect in the hydrogen-bonded network, with both H_3O_4^+ and H_5O_2^+ occurring only in the sense of 'limiting' or 'ideal' structures. The defect can become delocalized over several hydrogen bonds owing to quantum fluctuations. Solvent polarization induces a small barrier to proton transfer, which is washed out by zero-point motion. The proton can consequently be considered part of a 'low-barrier hydrogen bond'^{15,16}, in which tunnelling is negligible and the simplest concepts of transition-state theory do not apply. The rate of proton diffusion is determined by thermally induced hydrogen-bond breaking in the second solvation shell.

Simulating an excess proton in liquid bulk water has proved to be immensely challenging. Based on the efforts of numerous groups, many insights into the microscopic nature of proton hydration and diffusion have been deduced^{17–31,38}. Structural diffusion² as a dynamical process was first 'seen' in microscopic detail in an *ab initio* molecular dynamics study of D^+ in D_2O (ref. 23). It was seen that proton diffusion does not occur via hydrodynamic Stokes diffusion of a rigid complex, but via migration of a structural defect due to a continual interconversion between covalent and hydrogen bonds. Solvent fluctuations modulate the proton transfer barrier and preselect a migration path. The rate-limiting step is the fluctuation-induced breakage of a hydrogen bond between the first and second solvation shell of H_3O^+ , which reduces the coordination number of a water molecule in the first solvation shell²³. It was subsequently suggested that quantum effects could potentially be important for the rattling of the proton in the hydrogen bond²⁴, but it has recently been shown that this depends sensitively on a qualitatively correct model for the interactions²⁷.

The quantum-mechanical particle density 'snapshots' from the present simulations of the hydrated proton give a pictorial realization of the structural diffusion process. Initially, the defect is localized as an H_3O_4^+ structure possessing an H_3O^+ core that donates three hydrogen bonds to neighbouring water molecules (Fig. 1a). In the second frame (Fig. 1b), one of the three protons of the H_3O^+ core migrates along its hydrogen bond and forms an H_5O_2^+ complex, in which this proton becomes equally shared between two water molecules. As the transfer is completed, an H_3O_4^+ complex is formed once again, but now centred on a neighbouring core molecule (Fig. 1c). The onset of further migration is shown in Fig. 1d, where the defect converts into another H_5O_2^+ configuration. Overall (Fig. 1a–d), the structural defect is displaced over a distance corresponding to approximately twice the average water–water distance, that is, about 5 Å. Each individual particle, however, moves by only a fraction of an ångström.

The controversial details of this process can be revealed by examination of the two-dimensional distribution $P(\delta, R_{\text{OO}})$ of the displacement coordinate $\delta = R_{\text{O}_a\text{H}} - R_{\text{O}_b\text{H}}$ of a given proton relative to the instantaneous hydrogen-bond centre and the corresponding bond distance R_{OO} . (See Fig. 1 legend for nomenclature.) For the analysis to be as unbiased as possible, we begin by including all $\text{O}_a\text{H}_\text{O}_b$ triples, that is, all hydrogen bonds present in the periodic sample. This distribution (not shown) is characterized by two prominent peaks around $(\delta, R_{\text{OO}}) \approx (\pm 0.9, 2.8)$ Å arising from the hydrogen bonds of bulk water but is manifestly non-zero around $|\delta| \approx 0$ Å. This supports the existence of centrosymmetric complexes, in which the proton is equally shared between two water molecules, and thereby opposes a description solely in terms of H_3O^+ or H_3O_4^+ complexes and/or proton transfer entirely associated with tunnelling.

The analysis can be refined by excluding 'irrelevant' hydrogen bonds in a two-step procedure. First, the H_3O_4^+ defect site is located, and only its three $\text{O}_a\text{H}_\text{O}_b$ triples are taken into account. Then, from these three triples, the hydrogen bond with the smallest $|\delta|$ is selected, an intuitive choice based on events of the type in Fig. 1. This second step focuses on the 'most active proton', that is, the one

that has the greatest likelihood of transferring to a neighbouring water molecule, but imposes no *a priori* discrimination between different structures such as H_5O_2^+ or H_9O_4^+ . In the distribution from the first step (not shown), the centrosymmetric $|\delta| \approx 0 \text{ \AA}$ region is substantially enhanced, and two ‘wings’ emerge, which can be attributed to asymmetric hydrogen-bonded defects. After the second step, the distribution (Fig. 2a) acquires a broad and unstructured character. This important finding indicates that the protonic defect can assume many different structures, so that an unambiguous distinction between H_5O_2^+ and H_9O_4^+ can no longer be achieved.

To analyse ‘limiting’ forms of the defect, two windows in the displacement coordinate δ are chosen. Tight δ -margins are deliberately used so that the structural analysis is rendered as evident as possible (see Fig. 2 legend for details). For small $|\delta|$, the structures are found to correspond to an equal sharing of the excess proton between two intact H_2O , in accordance with Zundel’s H_5O_2^+ view of the defect. In contrast, configurations with large $|\delta|$ possess the characteristics of a threefold coordinated H_3O^+ ion, that is, Eigen’s

H_9O_4^+ complex. There are, however, configurations (according to the present δ -margins about half of them) that cannot be characterized as resembling either of these idealized forms.

The fact that a significant portion of the observed structures remain unclassified is also reflected in the shape of $P(\delta, R_{\text{OO}})$ in Fig. 2a. Its structureless ridge allows the defect to evolve easily from H_5O_2^+ to H_9O_4^+ regions, thus implying that the H_5O_2^+ complex cannot be regarded as a typical transition state. Rather, the proton in the most active hydrogen bond oscillates between two water molecules without appreciable hindrance, thereby interconverting two H_3O^+ cores on neighbouring sites via a strongly hydrogen-bonded, centrosymmetric H_5O_2^+ complex. The corresponding effective free-energy barrier (around $\delta = 0 \pm 0.05 \text{ \AA}$ and $R_{\text{OO}} \approx 2.46\text{--}2.48 \text{ \AA}$) is vanishingly small, $\leq 0.15 \text{ kcal mol}^{-1}$, only a fraction of the thermal energy $k_B T \approx 0.59 \text{ kcal mol}^{-1}$ at room temperature. Evidently, both H_5O_2^+ and H_9O_4^+ are only important as limiting structures, and numerous unclassifiable situations exist in between. Therefore, the protonic defect is best visualized as a ‘fluxional complex’.

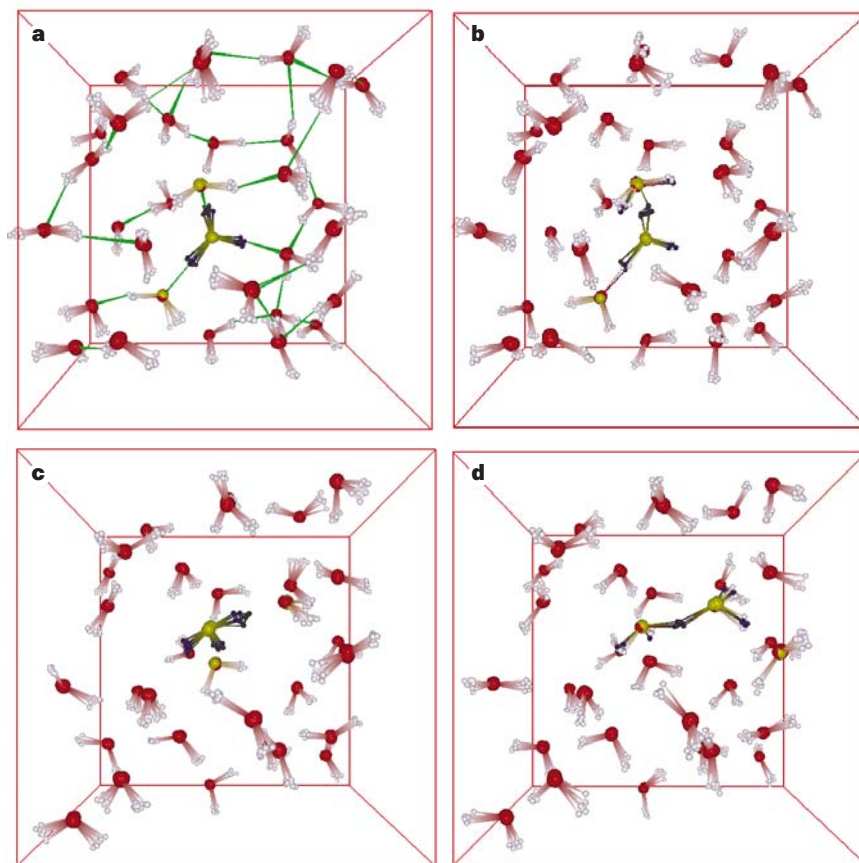


Figure 1 Examples of simulation configurations. We show a perspective view of the particle densities (diagonal part of the nuclear many-body density matrix) of representative single quantum configurations belonging to the proton migration path of the quantum trajectory. We note that the employed *ab initio* path integral molecular dynamics technique^{12–14} does not rigorously produce real-time evolution, but that event sequences can nevertheless be inferred qualitatively. Only the non-periodically replicated atoms (oxygens, red; hydrogens, grey) and bonds are included; hydrogen bonds (green) are sketched only in **a**, the blurring of the atom positions is exclusively due to quantum effects. The defect (given by the H_3O_2^+ core and the triple O_3H_3 defined in Fig. 2 legend) is indicated separately in every panel by yellow (oxygens) and black (hydrogens). The simulations were performed using CPMD Version 3.0 (J. Hutter, P. Ballone, M. Bernasconi, P. Focher, E. Fois, S. Goedecker, D. Marx, M. Parrinello & M. Tuckerman, Max-Planck-Institut für Festkörperforschung and IBM Zurich Research Laboratory 1995–96) at

ambient conditions using 32 H_2O with one excess proton in a periodic cubic box with a lattice constant of $9.8652 \text{ \AA}$, employing the BLYP density functional^{34,35}, Troullier–Martins pseudo potentials³⁶, and a Γ -point plane wave expansion of the valence orbitals up to 70 rydberg. This particular scheme has been shown to yield a very good description of the water dimer³⁷, the protonated water dimer¹⁶, and bulk water³⁷. The path integral for the quantum simulation^{12–14} was discretized using $P = 8$ Trotter replicas. The normal-mode propagator together with a chain of three coupled Nosé–Hoover thermostats on each degree of freedom was used to generate a canonical distribution at 300 K, improve sampling efficiency, and ensure ergodicity¹⁴. After careful equilibration, quantum and classical trajectories consisting of 100,000 and 160,000 configurations, respectively, were generated (using identical parameters, in particular the same Nosé–Hoover-chain thermostatting scheme).

Having proposed a coherent picture of the structural defect, it remains to identify the rate-determining step. This must be done indirectly, as the path integral approach used provides rigorously only time-averaged quantum equilibrium properties. Comparing the number of distinct net displacements of the defect (such as illustrated in Fig. 1) with the number of (inactive) proton-rattling events, the simulation unambiguously shows that the latter process is not rate-limiting. Rather, a proton localized as H_3O^+ migrates preferentially along the hydrogen bond to the neighbouring H_2O that has the lowest coordination number. A water molecule in the first solvation shell becomes undercoordinated by thermally induced hydrogen-bond breaking between it and a second solvation shell member²³. This phenomenon leads to the coordination number distribution $n(\delta, R_{\text{OO}})$ in the most active hydrogen bond, which is superimposed on $P(\delta, R_{\text{OO}})$ in Fig. 2a. Here, the proton-accepting water molecule is undercoordinated near $|\delta| \approx 0 \text{ \AA}$ (blue to green), where the defect resembles H_5O_2^+ , relative to the wings of the distribution at large $|\delta|$ (red to yellow), where the defect resembles H_9O_4^+ .

How well does a purely classical description of the nuclei, which includes only thermal fluctuations, compare with these quantum results? The answer can be gleaned from Fig. 2b and from the free-energy barrier. In the classical limit, the asymmetric H_9O_4^+ structure

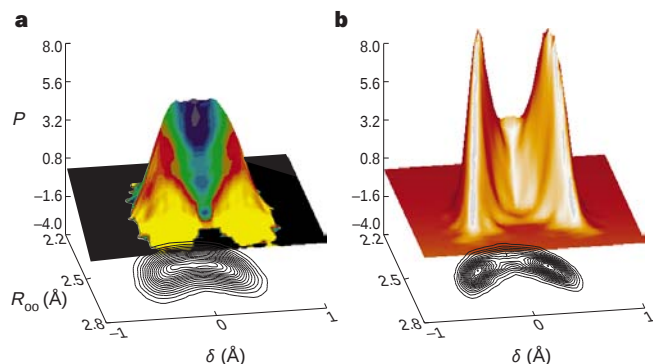


Figure 2 Representation of relative proton positions. Averaged distribution function $P(\delta, R_{\text{OO}})$ of the proton displacement $\delta = R_{\text{O}_a\text{H}} - R_{\text{O}_b\text{H}}$ relative to the $\text{O}_a\text{H}\text{O}_b$ bond midpoint (asymmetric stretch coordinate) and the corresponding oxygen-oxygen separation R_{OO} between the two neighbouring oxygen atoms O_a and O_b from quantum (a) and classical (b) simulations at 300 K. (Definition of the $\text{O}_a\text{H}\text{O}_b$ triple: O_a is the centre of the H_3O^+ unit determined by assigning every proton to its closest oxygen neighbour; H is one of the three protons forming the H_3O^+ unit, such that O_a , H and O_b yield the minimum value of $|\delta|$ in each configuration.) In a, the quantum averaged local coordination number, $n(\delta, R_{\text{OO}})$, of O_b is superimposed on $P(\delta, R_{\text{OO}})$. The function $n(\delta, R_{\text{OO}})$ is obtained by averaging the number of oxygen neighbours of O_b within a sphere of radius R^{cut} for given δ and R_{OO} ; $R^{\text{cut}} = 3.25 \text{ \AA}$ is determined by the first minimum of the oxygen-oxygen radial distribution function $g_{\text{OO}}(r)$ of bulk water. (Colour code: n decreases from yellow (~ 4) to red to green to blue (~ 3.5)). All distributions are smoothed for presentation and symmetrized about $\delta = 0 \text{ \AA}$. In addition, the $P(\delta, R_{\text{OO}})$ distributions are normalized to unity and shown on the same scale. Structural analysis of configurations: (1) for small $|\delta| \leq 0.1 \text{ \AA}$, the first peak of the intra-complex $g_{\text{OO}}(r)$ occurs at $r \approx 2.4 \text{ \AA}$ (compared with $2.8 \text{ \AA}$ for bulk water). The first peak of $g_{\text{OH}}(r)$ is around $0.95 \text{ \AA}$ (close to the value found in bulk water); a secondary peak occurs close to $1.25 \text{ \AA}$ (roughly half of the average OO distance), however, no clearly distinguishable feature exists at $r \approx 1.8 \text{ \AA}$ (where the first intermolecular peak for bulk water would be expected). This pattern for small $|\delta|$ is consistent with an $\text{H}_2\text{O} \cdots \text{H}^+ \cdots \text{OH}_2$ complex, in which a proton is equally shared between two normal water molecules. (2) For large $|\delta| > 0.3 \text{ \AA}$, $g_{\text{OO}}(r)$ peaks near $2.5 \text{ \AA}$, and the first peak of $g_{\text{OH}}(r)$ is, in this case, close to $1.0 \text{ \AA}$ (which is slightly longer than for H_2O molecules in the bulk). This corresponds to an $\text{H}_3\text{O}^+ \cdots (\text{H}_2\text{O})_3$ complex, where the H_3O^+ core molecule is roughly threefold coordinated. The first-shell H_2O molecules are themselves approximately fourfold coordinated.

is stabilized owing to an increase in the thermal barrier to $\sim 0.56 \text{ kcal mol}^{-1}$ (approximately $k_B T = 0.59 \text{ kcal mol}^{-1}$). Thus, it is the zero-point energy of the most active proton that shifts the lowest vibrational state towards the top of the classical barrier, thereby enhancing H_5O_2^+ -like complexes, so that the fluctuations of the surrounding water molecules become crucial. This idea is similar to the concept of 'adiabatic proton transfer'²⁵. Thus, mechanisms such as tunnelling through the classically existing barrier and/or soliton-mediated coherent transfer are of only minor importance. A similar 'low-barrier hydrogen bond' situation^{15,16} is also found in the early stage of the pressure-induced transition to symmetric ice X, where proton disorder along the hydrogen bonds is driven by zero-point motion (compare Fig. 2a with Fig. 3c of ref. 32). The essential difference is that in ice X, reduction of the oxygen-oxygen distances (and resulting low barriers) is caused by external pressure, whereas in water it is a consequence of the electrostrictive distortion of the hydrogen-bond network around the structural charge defect. Another indicator of the quantum nature of the structural defect is an increase in the radius of gyration of the H_3O^+ core (see Fig. 3 legend). The difference between the quantum and classical averages, $1.3 \text{ \AA}$ versus $0.9 \text{ \AA}$, is due entirely to quantum delocalization of the defect over several hydrogen bonds. The structure of such 'non-classical' configurations (see Fig. 3) suggests that these quantum delocalizations, in a sense, 'probe the most likely direction for further proton migration', thus enhancing the thermal fluctuations that mainly drive the structural diffusion.

The analysis performed here underlines the complex behaviour of the hydrated proton, bringing out features of both Eigen's and Zundel's views, but it also emphasizes that the defect cannot be fully understood using either of these. One tends to think of solution complexes in terms of well-defined structures with transition states in between. We have shown that this appears to be a misleading concept in the context of proton diffusion, and many of the difficulties in the interpretation of the experimental data might

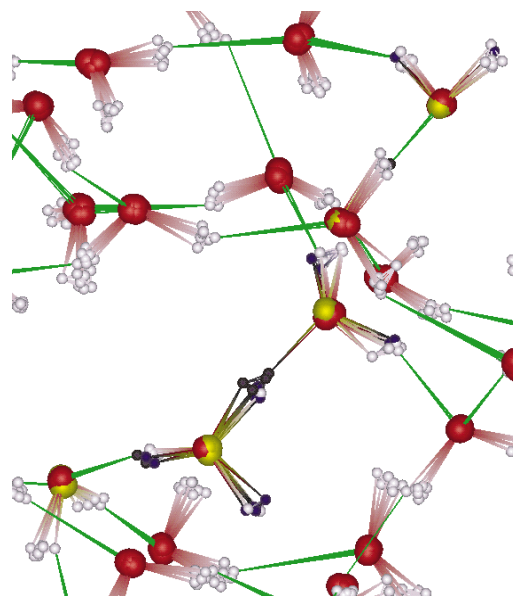


Figure 3 Particle density of a representative delocalized quantum configuration. (See Fig. 1 legend for the definition of the defect and the details of the presentation.) The quantum fluctuations lead to a structural defect that is delocalized over approximately four different hydrogen bonds (lower-left to the upper-right corner). The average 'size' of the defects can be quantified by the radius of gyration $R_{\text{H}_3\text{O}}^{\text{qr}} = \langle (1/NP) \sum_s \sum_i (\mathbf{R}_i^s - \mathbf{R}^s)^2 \rangle^{1/2}$. Here, $\mathbf{R}^s$ is the centroid of the defect given by $\mathbf{R}^s = (1/NP) \sum_s \sum_i \mathbf{R}_i^s$, i takes values of 1 to N (where $N = 4$, the number of atoms in the H_3O^+ core), $s = 1, \dots, P$ is the Trotter replica index (note that $P = 1$ classically), and angle brackets denote the canonical average.

arise from the attempt to interpret complex behaviour within the limits of a favoured structure. The main conclusions drawn from the present first-principles quantum simulations, together with those of ref. 23, are in accordance with a model that is deduced from, and is claimed to be consistent with, all known experimental data³³. □

Received 12 August; accepted 9 November 1998.

1. de Grotthuss, C. J. T. Sur la décomposition de l'eau et des corps qu'elle tient en dissolution à l'aide de l'électricité galvanique. *Ann. Chim.* **LVIII**, 54–74 (1806).
2. Atkins, P. W. *Physical Chemistry* 6th edn, Ch. 24.8, 741 (Oxford Univ. Press, 1998).
3. Hückel, E. Theorie der Beweglichkeiten des Wasserstoff- und Hydroxylions in wässriger Lösung. *Z. Elektrochem.* **34**, 546–562 (1928).
4. Stearn, A. E. & Eyring, J. The deduction of reaction mechanisms from the theory of absolute rates. *J. Chem. Phys.* **5**, 113–124 (1937).
5. Bernal, J. D. & Fowler, R. H. A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions. *J. Chem. Phys.* **1**, 515–548 (1933).
6. Wannier, G. Die Beweglichkeit des Wasserstoff- und Hydroxylions in wässriger Lösung. *Ann. Phys. (Leipzig)* **24**, 545–590 (1935).
7. Huggins, M. L. Hydrogen bridges in ice and liquid water. *J. Phys. Chem.* **40**, 723–731 (1936).
8. Wicke, E., Eigen, M. & Ackermann, Th. Über den Zustand des Protons (Hydroniumions) in wässriger Lösung. *Z. Phys. Chem. (N.F.)* **1**, 340–364 (1954).
9. Eigen, M. Proton transfer, acid–base catalysis and enzymatic hydrolysis. *Angew. Chem. Int. Edn Engl.* **3**, 1–19 (1964).
10. Zundel, G. & Metzger, H. Energiebänder der tunnelnden Überschuss-Protonen in flüssigen Säuren. Eine IR-spektroskopische Untersuchung der Natur der Gruppierungen H_3O_2^+ . *Z. Physik. Chem. (N.F.)* **58**, 225–245 (1968).
11. Zundel, G. in *The Hydrogen Bond—Recent Developments in Theory and Experiments. II. Structure and Spectroscopy* (eds Schuster, P., Zundel, G. & Sandorfy, C.) 683–766 (North-Holland, Amsterdam, 1976).
12. Marx, D. & Parrinello, M. Ab initio path-integral molecular dynamics. *Z. Phys. B (Rapid Note)* **95**, 143–144 (1994).
13. Marx, D. & Parrinello, M. Ab initio path integral molecular dynamics: basic ideas. *J. Chem. Phys.* **104**, 4077–4082 (1996).
14. Tuckerman, M. E., Marx, D., Klein, M. L. & Parrinello, M. Efficient and general algorithms for path integral Car–Parrinello molecular dynamics. *J. Chem. Phys.* **104**, 5579–5588 (1996).
15. Cleland, W. W. & Kreevoy, M. M. Low-barrier hydrogen bonds and enzymic catalysis. *Science* **264**, 1887–1890 (1994).
16. Tuckerman, M. E., Marx, D., Klein, M. L. & Parrinello, M. On the quantum nature of the shared proton in hydrogen bonds. *Science* **275**, 817–820 (1997).
17. Guissani, Y., Guillot, B. & Bratos, S. The statistical mechanics of the ionic equilibrium of water: a computer simulation study. *J. Chem. Phys.* **88**, 5850–5856 (1988).
18. Halley, J. W., Rustad, J. R. & Rahman, A. A polarizable, dissociating molecular dynamics model for liquid water. *J. Chem. Phys.* **98**, 4110–4119 (1993).
19. Tuñón, I., Silla, E. & Bertrán, J. Proton solvation in liquid water. An ab initio study using the continuum model. *J. Phys. Chem.* **97**, 5547–5552 (1993).
20. Laria, D., Cicciotti, G., Ferrario, M. & Kapral, R. Activation free energy for proton transfer in solution. *Chem. Phys.* **180**, 181–189 (1994).
21. Komatsuzaki, T. & Ohmine, I. Energetics of proton transfer in liquid water. I. Ab initio study for origin of many-body interaction and potential energy surfaces. *Chem. Phys.* **180**, 239–269 (1994).
22. Wei, D. & Salahub, D. R. Hydrated proton clusters and solvent effects on the proton transfer barrier: a density functional study. *J. Chem. Phys.* **101**, 7633–7643 (1994).
23. Tuckerman, M., Laasonen, K., Sprik, M. & Parrinello, M. Ab initio molecular dynamics simulation of the solvation and transport of H_3O^+ and OH^- ions in water. *J. Phys. Chem.* **99**, 5749–5752 (1995); Ab initio molecular dynamics simulation of the solvation and transport of hydronium and hydroxyl ions in water. *J. Chem. Phys.* **103**, 150–161 (1995).
24. Lobaugh, J. & Voth, G. A. The quantum dynamics of an excess proton in water. *J. Chem. Phys.* **104**, 2056–2069 (1996).
25. Ando, K. & Hynes, J. T. Molecular mechanism of HCl acid ionization in water: ab initio potential energy surfaces and Monte Carlo simulations. *J. Phys. Chem. B* **101**, 10464–10478 (1997).
26. Schmidt, R. G. & Brickmann, J. Molecular dynamics simulation of the proton transport in water. *Ber. Bunsenges. Phys. Chem.* **101**, 1816–1827 (1997).
27. Sagnella, D. E. & Tuckerman, M. E. An empirical valence bond model for proton transport in water. *J. Chem. Phys.* **108**, 2073–2083 (1998).
28. Vuilleumier, R. & Borgis, D. Quantum dynamics of an excess proton in water using an extended empirical valence-bond hamiltonian. *J. Phys. Chem. B* **102**, 4261–4264 (1998).
29. Schmitt, U. W. & Voth, G. A. Multistate empirical valence bond model for proton transport in water. *J. Phys. Chem. B* **102**, 5547–5551 (1998).
30. Billeter, S. R. & van Gunsteren, W. F. Protonizable water model for quantum dynamical simulations. *J. Phys. Chem. A* **102**, 4669–4678 (1998).
31. Kochanski, E., Kelterbaum, R., Klein, S., Rohmer, M. M. & Rahmouni, A. Decades of theoretical work on protonated hydrates. *Adv. Quantum Chem.* **28**, 273–291 (1997).
32. Benoit, M., Marx, D. & Parrinello, M. Tunneling and zero-point motion in high-pressure ice. *Nature* **392**, 258–261 (1998).
33. Agmon, N. The Grotthuss mechanism. *Chem. Phys. Lett.* **244**, 456–462 (1995); Hydrogen bonds, water rotation and proton mobility. *J. Chim. Phys. Phys.-Chim. Biol.* **93**, 1714–1736 (1996).
34. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).
35. Lee, C., Wang, W. & Parr, R. G. Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785–789 (1988).
36. Troullier, N. & Martins, J. L. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B* **43**, 1993–2006 (1991).
37. Sprik, M., Hutter, J. & Parrinello, M. Ab initio molecular dynamics simulation of liquid water: comparison of three gradient-corrected density functionals. *J. Chem. Phys.* **105**, 1142–1152 (1996).
38. Ojamäe, L., Shavitt, I. & Singer, S. J. Potential models for simulations of the solvated proton in water. *J. Chem. Phys.* **109**, 5547–5564 (1998).

Acknowledgements. We thank K.-D. Kreuer for discussions. The calculations were performed on the Cray-T3E/816 of the Max-Planck-Gesellschaft.

Correspondence and requests for materials should be addressed to D.M. (e-mail: marx@pr.mpi-stuttgart.mpg.de).

Discrete alternating hotspot islands formed by interaction of magma transport and lithospheric flexure

Christoph F. Hieronymus* & David Bercovici†

* Danish Lithosphere Centre, Øster Voldgade 10 L, 1350 Copenhagen K, Denmark

† Department of Geology and Geophysics, University of Hawaii, Honolulu, Hawaii 96822, USA

The large-scale geometry and age progression of many hotspot island chains, such as the Hawaiian–Emperor chain, are well explained by the steady movement of tectonic plates over stationary hotspots. But on a smaller scale, hotspot tracks are composed of discrete volcanic islands whose spacing correlates with lithospheric thickness¹. Moreover, the volcanic shields themselves are often not positioned along single lines, but in more complicated patterns, such as the dual line known as the Kea and Loa trends of the Hawaiian islands^{2,3}. Here we make use of the hypothesis that island spacing is controlled by lithospheric flexure¹ to develop a simple nonlinear model coupling magma flow, which feeds volcanic growth, to the flexure caused by volcanic loads on the underlying plate. For a steady source of melt underneath a moving lithospheric plate, magma is found to reach the surface and build a chain of separate volcanic edifices with realistic spacing. If a volcano is introduced away from the axis of the chain, as might occur following a change in the direction of plate motion, the model perpetuates the asymmetry for long distances and times, thereby producing an alternating series of edifices similar to that observed in the Kea and Loa trends of the Hawaiian island chain.

Various models have been proposed to explain the formation of discrete islands within hotspot tracks. Periodicity in island formation may originate in the mantle by solitary waves^{4,5} on, or shear instabilities^{6,7} of, mantle plumes; these models, however, do not explain the observed correlation¹ between island spacing λ and lithospheric thickness H , that is, $\lambda \propto H^{3/4}$. Alternatively, stresses in the lithosphere may generate joints with regular spacing of the order of the thickness of the lithosphere with volcanic eruptions occurring at their intersections⁸; however, the stresses necessary to cause the observed pattern of volcanoes are not well constrained and may require multiple, somewhat *ad hoc*, origins⁸. Our model instead builds on the idea (motivated by the observed correlation between λ and H) that magma flow occurs at locations where the horizontal flexural stresses due to existing islands are approximately zero¹ (that is, at the transition between the flexural depression and the flexural bulge). The model provides a simplified yet self-consistent dynamical system that generates realistic results by taking into account the effects of flexural stresses and erosion of the dyke walls on fracture formation.

The physical processes involved in our model act on greatly varying length and time scales. Whereas the flexural wavelength is of the order of 100 km (ref. 9), and the flexural amplitude changes over timescales of volcano formation ($\sim 10^6$ years; ref. 10), a typical magma-transporting fracture is only 1–10 m wide and forms over timescales of days (ref. 11). A practical approach to treating this disparity in scales is to define the lithosphere's permeability to magma percolation as a continuous field variable, thus treating the average effects of fractures in a continuum mechanical sense. By analogy with the physics of fracture formation and evolution, the permeability is assumed to depend on stresses^{11,12} and total erosion (for example, melting^{13,14} and thermomechanical erosion¹⁵) of the fracture walls.

We first assume that the stresses influencing the permeability are due to both an impinging mantle plume, and lithospheric flexure (Fig. 1). The net effect of the plume on the base of the lithosphere is represented as a tensile stress or over-pressurization (due to inflation of the hotspot swell, horizontal spreading of the plume top, and/or high pressure of the melt region) of super-gaussian¹⁶ shape placed in the frame of reference of a lithosphere moving with velocity v in the negative x -direction:

$$\sigma_p(x, y, t) = Ae^{-\left[\frac{(x-vt)^2+y^2}{r^2}\right]^p} \quad (1)$$

Here A is the maximum stress, and r is the half-width of the source; p provides generality by allowing σ_p any shape, from purely gaussian ($p = 1$) to plateau-shaped ($p \gg 1$).

The stresses associated with plate flexure are assumed to be primarily a function of the surface load as described in thin-plate theory; the vertical deflection w caused by a load of height h (which is variable in space and time) and density ρ_m is therefore given by

$$D\nabla_h^4 w + (\rho_a - \rho_w)gw = -\rho_m gh \quad (2)$$

where D is the flexural rigidity⁹, ρ_a and ρ_w are the densities of the asthenosphere and water, respectively, and ∇_h^4 is the two-dimensional horizontal biharmonic operator. In equation (2), the effects of the basal load are assumed to be negligible relative to those due to the volcanic load. The permeability of the lithosphere itself is primarily controlled by the sum of the horizontal normal flexural stresses near the surface (where fracturing is most likely to occur¹⁷) given by

$$\sigma_f = -\frac{6D}{H^2}\nabla_h^2 w \quad (3)$$

where H is the elastic lithospheric thickness⁹. (For nearly vertical fractures, only the horizontal stresses need to be considered and hence the vertical compression due to either the plume or the volcanic load is neglected.) Thus, for example, a net compressive stress ($\sigma_f < 0$) acts like an excess confining pressure on magma pathways.

The influence of erosion of fracture walls (due to a combination of melting, thermomechanical erosion, hydrating chemical reactions^{18,19}, and so on) on permeability is controlled by the rate of influx and efflux of net energy due to magma injection. This energy nominally includes all forms of energy that contribute to erosion, for example, heat (for melting and for thermal condition-

ing which facilitates erosion by softening the rock), gravitational potential energy, internal mechanical energy, kinetic energy (for mechanical erosion), and chemical potential energy (for reactive or hydrating effects). The various erosional effects and their interactions are difficult to estimate. Although evidence for melting of wall rock is found in Hawaiian lavas²⁰, it is unlikely to provide the necessary erosion by itself (based on silicate latent heats and estimates of magma heat loss). The loss of gravitational potential energy of magma provides a large source of power for mechanical erosion for which there is evidence in the presence of xenoliths²¹; however, the effectiveness of such erosion is difficult to estimate as it depends on the friability of wall rock, the geometry (such as tortuosity) of the pathways, and many other parameters. However, our model suggests that the formation of discrete islands requires significant net erosion to allow established pathways to remain open in the presence of mounting compressive stresses. Thus, while we cannot quantitatively determine the net erosional effect based on observations and experiments, we do find, *a posteriori*, that the effect must be significant to permit island formation in this model, a prediction that clearly warrants testing with future field and laboratory studies.

Overall, we assume that stress and erosion affect permeability by changing the magma volume fraction (that is, porosity) f due to dykes and cracks. Incremental changes in f are thus assumed to be controlled by increments in net energy and normal stresses

$$Cd f = B d\epsilon + d\sigma_p + d\sigma_f \quad (4)$$

where ϵ is the net energy content of the bulk medium (wall rock plus magma) per volume in excess of the far-field value, B is a constant, and C is a hybrid parameter whose inverse C^{-1} represents the responsiveness of crack or dyke volume to stress and erosion (that is, a large value of C causes a small change in f for a given input of energy or stress). We assume that, to first order, $C = C_0/(f_{\max} - f)$ which precludes unphysical values of $f < 0$; $f_{\max}$ is the value of f beyond which increases in stress and net energy are partitioned less into changing porosity than into other responses, such as heating or deforming the magma. We assume that this limit is similar to the disaggregation limit for melts, that is, $0.01 \leq f_{\max} \leq 0.1$ (however, $f_{\max}$ need not be evaluated explicitly in this model as it only appears coupled to other parameters). Equation (4) is easily integrated to yield the function $f(\epsilon, \sigma_p, \sigma_f)$. Then, assuming that permeability k is linearly dependent on f , and that, for simplicity, magma percolation obeys Darcy's law, the magma volume flux per area (the Darcy velocity) is given by

$$V = \frac{k_0 f_{\max}}{2\mu} \left[1 + \tanh\left(\frac{\sigma_p + \sigma_f - \sigma_0 + B\epsilon}{2C_0/f_{\max}}\right) \right] \times \left[(\rho_l - \rho_m)g - (\rho_m h + \rho_a w) \frac{g}{H} \right] \quad (5)$$

where k_0 is a reference permeability, μ is the magma viscosity, σ_0 is an integration constant representing the intrinsic tensile strength of the far-field lithosphere¹², and ρ_l is lithospheric density. The pressure gradient driving flow (the last bracketed factor in equation (5)) is due to the magma buoyancy, the hydraulic head of magma inside the volcano, and the flexural deflection of lithosphere into the underlying asthenosphere. (If σ_p represents magma overpressurization instead of plume tensile stress, then, strictly speaking, it should be included in the driving pressure; however, this is found to have little significant effect on the final results.)

The volcanic load height h grows according to the ejection rate of magma onto the surface, V ; changes in ϵ are primarily controlled by the injection rate of magmatic net energy, which is also proportional to V . Thus we may write

$$\frac{\partial h}{\partial t} = V = \frac{H}{\Delta Q} \frac{\partial \epsilon}{\partial t} \quad (6)$$

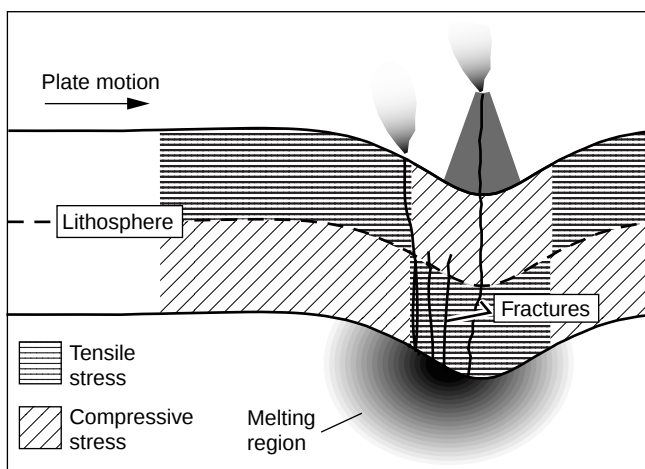


Figure 1 Sketch of the influence of flexural stresses on fracture formation. Fractures form where the magma pressure due to the plume is greater than the sum of compressive stresses due to flexure and the tensile strength of the lithosphere. Once a magma pathway is established, the fracture is kept open in an environment of increasing flexural stresses by erosion of the fracture walls.

where ΔQ is the drop in magmatic energy per volume across the lithosphere (from bottom to top). This final equation provides, first, a description of how h changes with time, t , and second the relation $\epsilon = \Delta Qh/H$ (assuming ϵ and h are both zero at some initial time) which we can then substitute back into equation (5) to eliminate ϵ . The resulting governing equations (1), (2), (3), (5) and (6) with the aforementioned substitutions describe how the rate of change of volcano height h is a strongly nonlinear function of h itself.

Numerical solutions of the governing equations yield chains of discrete volcanoes (Fig. 2) for a wide range of parameters. The first magma pathways form in pristine, unflexed lithosphere, after which a volcanic load begins to grow. The mounting compressive flexural stresses (σ_f) beneath the volcanic load soon overwhelm the tensile stresses (σ_p) from the plume magma source, thereby impeding and eventually precluding the initiation of any further magma pathways in the vicinity of this volcano. However, as the lithosphere moves on, the plume magma source eventually reaches a region of the lithosphere where the flexural compressive stresses are low enough that it can force open new magma pathways to the surface. Erosion provides a positive feedback which keeps established pathways open long enough to form a volcano, even as the compressive flexural stresses grow along with the volcanic load. (The flow in the

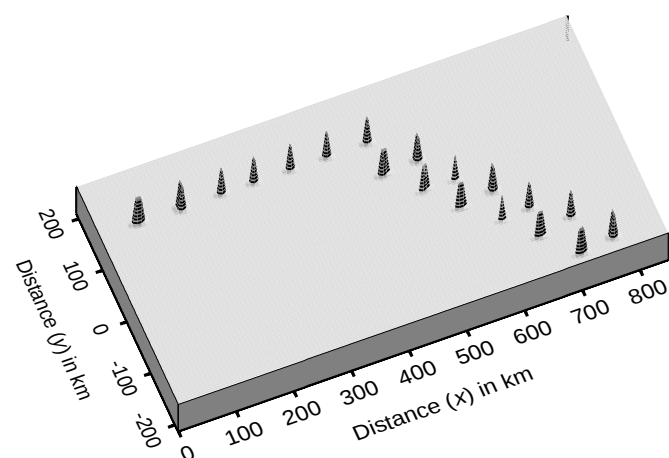


Figure 2 Theoretical model of discrete islands formed on a lithospheric plate moving over a mantle hotspot. In the lithospheric reference frame, the hotspot starts at $x = 70$ km, $y = 110$ km and moves to the right at velocity $v = 10$ cm yr $^{-1}$, causing a line of single volcanoes. When the hotspot reaches $x = 470$ km, the relative direction of plate motion changes by 45°, causing a transition to a dual line of volcanoes (see text for discussion). The governing equations are solved using a pseudo-spectral code²⁷ with 256 × 128 Fourier modes. The volcanoes are relatively sharp peaks rather than proper shields because gravity-induced spreading at the surface is neglected. The parameters used to generate the figure are: $D = 2.38 \times 10^{23}$ N m (ref. 9), $H = 30$ km, $\rho_w = 1,000$ kg m $^{-3}$, $\rho_a = 3,300$ kg m $^{-3}$, $\rho_m = 2,600$ kg m $^{-3}$, $\rho_l = 3,400$ kg m $^{-3}$, $r = 110$ km, $\mu = 100$ Pa s (ref. 13), and $\sigma_o = 5.1 \times 10^7$ Pa (ref. 28). (See text for definitions of symbols.) We choose $p = 3$, which results in a plateau-shaped σ_p that facilitates generation of the dual line by providing a wider swath of nearly equal probability of volcano formation; a more peaked function tends to concentrate volcanic edifices on the hotspot axis. Parameter $k_o f_{\max} = 1.8 \times 10^{-11}$ m 2 is estimated using $F/A_v \approx k_o f_{\max}/\mu(\rho_l - \rho_m)g$, where F is the volume flux of magma (~ 3.6 m 3 s $^{-1}$ for Kilauea volcano²⁹), and A_v is the volcano area ($\sim 2,500$ km 2). We choose $A = 5.0 \times 10^7$ Pa, resulting in plume tensile stress or magma pressure that is of the order of the tensile strength σ_o ; and $C_o f_{\max} = 1.8 \times 10^5$ Pa so that, to be conservative, the minimum resistance to changes in porosity C is also of the order of the tensile strength for plausible $f_{\max}$. The value of $B\Delta Q = 6.3 \times 10^9$ Pa causes the effect of wall-rock erosion to be comparable to or even larger than the flexural stresses, as is a posteriori found to be required for the formation of discrete islands.

established pathways is eventually shut off by the hydraulic head of magma within the volcano itself.) Once the next volcano is established, the process repeats itself. The resulting volcanic spacing λ is proportional to $H^{3/4}$ because the natural length scale of the flexure equation (2) is proportional to $D^{1/4}$ and $D \propto H^3$ (ref. 9). With lithospheric thickness $H = 30$ km (refs 22, 23), we can match the observed volcano spacing of 70 km in the Hawaiian islands¹.

Our simple model predicts that a dual line of volcanoes similar to the Kea and Loa trends of the Hawaiian chain, and comparable trends in other Pacific island chains²⁴, is a natural consequence of the dynamics of the system when it is initiated or perturbed by a volcanic load placed off the hotspot axis, as might occur after a change in tectonic-plate motion. The dual chain occurs because the compressive stresses beneath the off-axis edifice preclude formation of the next volcanoes on that side of the axis; thus, the closest region susceptible to new volcano formation is on the opposite side of the hotspot axis. Subsequent volcanoes form similarly, leading to a chain reaction that results in alternating positions of volcanoes. The eventual separation of the two lines depends on the radius r of the plume magma source; a separation characteristic of the Kea and Loa trends is achieved for $r = 110$ km. This value of r is comparable with estimates of the melt radius (100–150 km) of the Hawaiian plume^{2,8}, which suggests that the plume stress is caused more by a high-pressure melt region than by other sources of plume stress, such as inflation of the hotspot swell (which has a much larger radius¹⁶). However, the spacing between nearest volcanoes is the same as that for the single line, about 70 km. Moreover, the two volcanic lines tend to sample different parts of the plume magma source; this may help explain the disparate geochemistry and petrology of volcanoes on single hotspots as they would arise from regions with different temperatures, degrees of partial melting, and possibly entrainment²⁵.

The present Kea and Loa trends are predicted by our model to be due to an off-axis asymmetry caused by, and perpetuated since, a change in direction of plate motion. As these trends are relatively young, or at least have been strongest in the past 4 Myr, they may reflect a change in plate motion that occurred not 43 Myr ago (the time of the Emperor–Hawaiian bend) but, as recently suggested²⁶ only 3–6 Myr ago. □

Received 8 September; accepted 11 December 1998.

1. Ten Brink, U. Volcano spacing and plate rigidity. *Geology* **19**, 397–400 (1991).
2. Jackson, D. E., Silver, E. A. & Dalrymple, G. B. Hawaiian–Emperor chain and its relation to Cenozoic circum-Pacific tectonics. *Geol. Soc. Am. Bull.* **83**, 601–618 (1972).
3. Tatsumoto, M. Isotopic composition of lead in oceanic basalts and its implications to mantle evolution. *Earth Planet. Sci. Lett.* **38**, 63–87 (1978).
4. Olson, P. in *Magma Transport and Storage* (ed. Ryan, M. P.) 33–51 (Wiley, New York, 1990).
5. Schubert, G., Olson, P., Anderson, C. A. & Goldman, P. Solitary waves in mantle plumes. *J. Geophys. Res.* **94**, 9523–9532 (1989).
6. Skilbeck, J. N. & Whitehead, J. A. Formation of discrete islands in linear island chains. *Nature* **272**, 499–501 (1978).
7. Ihinger, P. D. Mantle flow beneath the Pacific plate; evidence from seamount segments in the Hawaiian–Emperor chain. *Am. J. Sci.* **295**, 1035–1057 (1995).
8. Vogt, P. R. Volcano spacing, fractures, and thickness of the lithosphere. *Earth Planet. Sci. Lett.* **21**, 235–252 (1974).
9. Turcotte, D. L. & Schubert, G. *Geodynamics* (Wiley, New York, 1982).
10. Guillou, H., Garcia, M. O. & Turpin, L. Unspiked K–Ar dating of young volcanic rocks from Loihi and Pitcairn hot spot seamounts. *J. Volcan. Geotherm. Res.* **78**, 239–249 (1997).
11. Lister, J. R. & Kerr, R. C. Fluid-mechanical models of crack propagation and their application to magma transport in dykes. *J. Geophys. Res.* **96**, 10049–10077 (1991).
12. Rubin, A. M. Propagation of magma filled cracks. *Annu. Rev. Earth Planet. Sci.* **23**, 287–336 (1995).
13. Bruce, P. M. & Huppert, H. E. in *Magma Transport and Storage* (ed. Ryan, M. P.) 87–101 (Wiley, New York, 1990).
14. Lister, J. R. & Dellar, P. J. Solidification of pressure-driven flow in a finite rigid channel with application to volcanic eruptions. *J. Fluid Mech.* **323**, 267–283 (1996).
15. Davies, G. F. Thermomechanical erosion of the lithosphere by mantle plumes. *J. Geophys. Res.* **99**, 15709–15722 (1994).
16. Wessel, P. Observational constraints on models of the Hawaiian hot spot swell. *J. Geophys. Res.* **98**, 16095–16104 (1993).
17. Kohlstedt, D. L., Evans, B. & Mackwell, S. J. Strength of the lithosphere: Constraints imposed by laboratory experiments. *J. Geophys. Res.* **100**, 17587–17602 (1995).
18. Bonatti, E. Not so hot “hot spots” in the oceanic mantle. *Science* **250**, 107–111 (1990).
19. Burnham, C. W. in *The Evolution of the Igneous Rocks* (ed. Yoder, H. S.) 439–482 (Princeton Univ. Press, 1979).
20. Eiler, J. M., Farley, K. A., Valley, J. W., Hofmann, A. W. & Stolper, E. M. Oxygen isotope constraints on the sources of Hawaiian volcanism. *Earth Planet. Sci. Lett.* **144**, 453–468 (1996).
21. Garcia, M. O. Petrography and olivine and glass chemistry of lavas from the Hawaii Scientific Drilling Project. *J. Geophys. Res.* **101**, 11701–11713 (1996).

22. Watts, A. B. An analysis of isostasy in the world's oceans: 1. Hawaiian-Emperor seamount chain. *J. Geophys. Res.* **83**, 5989–6004 (1978).
23. Wessel, P. A reexamination of the flexural deformation beneath the Hawaiian swell. *J. Geophys. Res.* **98**, 12177–12190 (1993).
24. Jackson, E. D. & Shaw, H. R. Stress fields in central portions of the Pacific plate: Delineated in time by linear volcanic chains. *J. Geophys. Res.* **80**, 1861–1874 (1975).
25. Hauri, E. H. Major-element variability in the Hawaiian mantle plume. *Nature* **382**, 415–419 (1996).
26. Wessel, P. & Kroenke, L. A geometric technique for relocating hotspots and refining absolute plate motions. *Nature* **387**, 365–369 (1997).
27. Canuto, C., Hussaini, M. Y., Quarteroni, A. & Zhang, T. A. *Spectral Methods in Fluid Dynamics* (Springer, New York, 1988).
28. Jaeger, J. C. & Cook, N. G. W. *Fundamentals of Rock Mechanics* 3rd edn 190–191 (Chapman & Hall, London, 1979).
29. Swanson, D. A. Magma supply rate at Kilauea volcano, 1952–1971. *Science* **175**, 169–170 (1972).

Acknowledgements. We thank N. Ribe, A. Rubin and Y. Ricard for helpful comments. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to D.B. (e-mail: dberco@soest.hawaii.edu).

A primitive fossil fish sheds light on the origin of bony fishes

Min Zhu*, Xiaobo Yu† & Philippe Janvier‡

* Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, PO Box 643, Beijing 100044, China

† Department of Biological Sciences, Kean University, Union, New Jersey 07083, USA

‡ URA12, Laboratoire de Paléontologie, Muséum National d'Histoire Naturelle, 8 rue Buffon, 75005 Paris, France

Living gnathostomes (jawed vertebrates) include chondrichthyans (sharks, rays and chimaeras) and osteichthyans or bony fishes. Living osteichthyans are divided into two lineages, namely actinopterygians (bichirs, sturgeons, gars, bowfins and teleosts) and sarcopterygians (coelacanths, lungfishes and tetrapods). It remains unclear how the two osteichthyan lineages acquired their respective characters and how their common osteichthyan ancestor arose from non-osteichthyan gnathostome groups^{1,2}. Here we present the first tentative reconstruction of a 400-million-year-old fossil fish from China (Fig. 1); this fossil fish combines features of sarcopterygians and actinopterygians and yet possesses large, paired fin spines previously found only in two extinct gnathostome groups (placoderms and acanthodians). This early bony fish provides a morphological link between osteichthyans and non-osteichthyan groups. It changes the polarity of many characters used at present in reconstructing osteichthyan inter-relationships and offers new insights into the origin and evolution of osteichthyans.

The fossil fish *Psarolepis romeri*³ was described on the basis of skull and lower jaw materials from the Lower Devonian strata (about 400 million years (Myr) BP) of Qujing, Yunnan, China. Materials assignable to the same genus from the Upper Silurian (about 410 Myr BP) of China⁴ and Vietnam⁵ made *Psarolepis* one of the earliest osteichthyans known so far. The skulls and lower jaws exhibit the overall morphology of sarcopterygians but also show characters found in primitive actinopterygians, such as a tooth-bearing median rostral and a lower jaw with five coronoids (rather than three as in most sarcopterygians)^{3,6}. *Psarolepis* was first placed within sarcopterygians, as a basal member of Dipniformes³ or among the basal members of Crossopterygii⁴. The new features revealed by the shoulder girdle and cheek materials reported here (Figs 1d, e, 2, 3) indicate that *Psarolepis* may occupy a more basal position in osteichthyan phylogeny.

Most *Psarolepis* specimens derive from four beds at the same locality in Qujing, the first bed being in the Yulongsi Formation (Pridoli), the second and third beds in the Xishancun Formation (early Lochkovian), and the fourth bed in the Xitun Formation (late

Lochkovian). The specimens described here are from the second and third beds. The shoulder girdles and cheek plates from the third bed are often preserved as internal moulds (Fig. 2a) or as external moulds showing the internal casts of sensory canals and the pore-canal system (Figs 2b, c, 3a). Like elements previously assigned to *Psarolepis*^{3–5}, these materials exhibit the unique ornamentation with large, closely spaced pores on the cosmine surface. This unique ornamentation forms the basis of their assignment to *Psarolepis*, which is also supported by other histological and morphological features and by their association with elements that are directly comparable to the type specimen.

The shoulder girdle (Fig. 2) bears a huge pectoral spine that extends posteriorly from a conspicuous ridge between the ventral and vertical laminae of the cleithrum, resembling the condition in some placoderms and acanthodians. In addition, the symmetrical fin spine⁴ indicates that *Psarolepis* may possess median spines in front of the unpaired fins (Fig. 1a, c), as in sharks and acanthodians. The median fin spine exhibits the same ornamentation as the associated parietal shield and lower jaw⁴; other taxa from the same bed (petalichthyids and galeaspid) cannot possess a similar ornamentation. Such paired and unpaired fin spines are unknown in early osteichthyans, whether sarcopterygians or actinopterygians¹. The only possible exceptions are two questionable Silurian forms, *Lophosteus*, which has indeterminable spine-like fragments⁷, and *Andreolepis*, which has a lateral projection of the cleithrum^{8,9}.

A critical issue is whether *Psarolepis* has the typical sarcopterygian monobasal articulation of the paired fins. Although the internal mould of the endoskeletal shoulder girdle (Fig. 2a) does not show

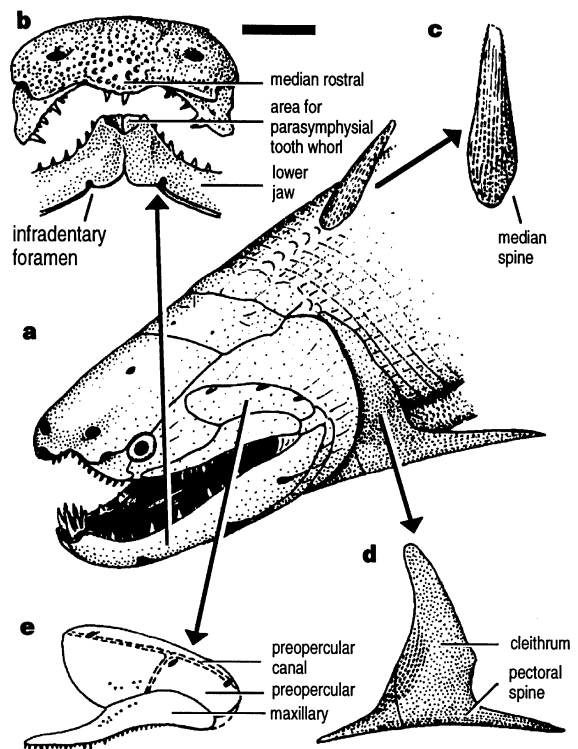


Figure 1 Reconstruction of *Psarolepis*, a 400-million-year-old sarcopterygian-like fish with an unusual combination of osteichthyan and non-osteichthyan features. **a**, Head and anterior part of the fish with tentatively positioned median fin spine. **b**, Anterior view of the skull and lower jaws (from ref. 3). Scale bar, 5 mm. **c**, Median fin spine (from ref. 4). **d**, Shoulder girdle with pectoral spine, based on specimens shown in Fig. 2. **e**, Cheek plate with maxillary and preopercular, based on specimens shown in Fig. 3. Surface ornamentation of the cheek plate is omitted to show the pattern of sensory canals. Most *Psarolepis* specimens derive from four beds at the same locality in Qujing, Yunnan, China.

the articulation surface of the fin skeleton clearly, its overall shape is strikingly different from that of both sarcopterygians and actinopterygians. It is a massive, plate-shaped bone pierced internally by several openings for blood vessels and nerves, in many ways like the condition in placoderms. The anterior part of the cleithrum has a denticulate postbranchial lamina (Fig. 2b), as in placoderms and actinopterygians. In contrast, the dorsal part of the cleithrum is high and pointed, as in actinopterygians, onychodonts and primitive coelacanth; it is different from the broad dorsal apex found in lungfishes, porolepiforms and osteolepiforms.

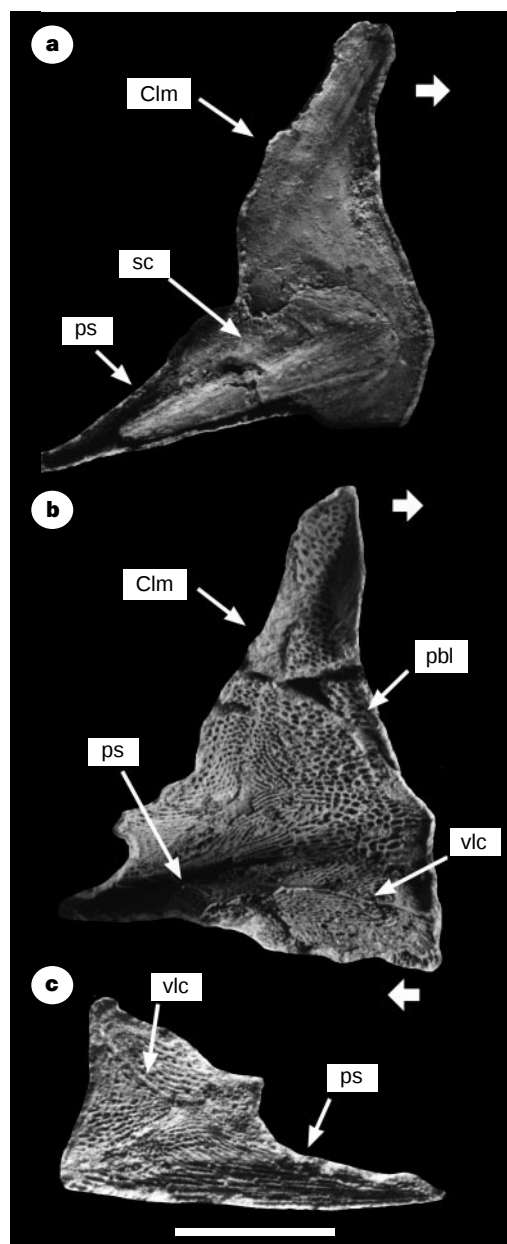


Figure 2 Shoulder girdles of *Psarolepis*. **a**, Specimen IVPP V11256.1: internal mould of a right cleithrum with pectoral spine (ps) and scapulocoracoid (sc). **b**, Specimen IVPP V11256.3: external mould of a left cleithrum with pectoral spine. **c**, Specimen IVPP V11256.4: external mould of an incomplete left cleithrum with pectoral spine. All specimens derive from the third bed in the Xishancun Formation (early Lochkovian) and are associated with undescribed cranial and lower jaw elements directly comparable to the type specimen. Short, thick arrows point to the anterior end of the fish. Scale bar, 10 mm. See Fig. 1d for a simplified reconstruction. Clm, cleithrum; pbl, postbranchial lamina; vlc, internal cast of canal on ventral lamina of cleithrum.

The external mould of the cheek plate from the third bed (Fig. 3a) matches the shape of a large, tilted preopercular in the cheek plate from the second bed (Fig. 3b); this latter cheek plate also shows a posteriorly expanded maxillary. The preopercular lacks a jugal canal (present in placoderms, acanthodians and chondrichthyans) but has a complete preopercular canal running along the dorsal margin of the bone (as in chondrichthyans, acanthodians and actinopterygians). Near the midpoint of the preopercular canal, a short vertical canal extends ventrally to the ventral margin of the preopercular. The dorsal portion of the preopercular carries three large foramina, similar to those found in the dermal cheeks of *Youngolepis*¹⁰ and *Kenichthys*¹¹ as well as in the lower jaws of *Psarolepis*^{3,4}, *Youngolepis*¹⁰, *Powichthys*¹² and some porolepiforms¹³. In addition to actinopterygians, sarcopterygian onychodonts¹³ also have a posteriorly expanded maxillary in broad contact with the preopercular. However, the presence of a complete preopercular canal, the lack of a jugal canal and the absence of a separate squamosal distinguish *Psarolepis* from all previously known sarcopterygians.

To examine the phylogenetic position of *Psarolepis* and its impact on osteichthyan relationships, we used the data set in ref. 14 to construct an expanded matrix with 37 taxa and 149 characters (see Methods). As the codings for four characters in ref. 14 were modified by data in ref. 4, we used the phylogenetic package PAUP¹⁵ to analyse the expanded matrix twice, first with the modified codings from ref. 4, then with the original codings from

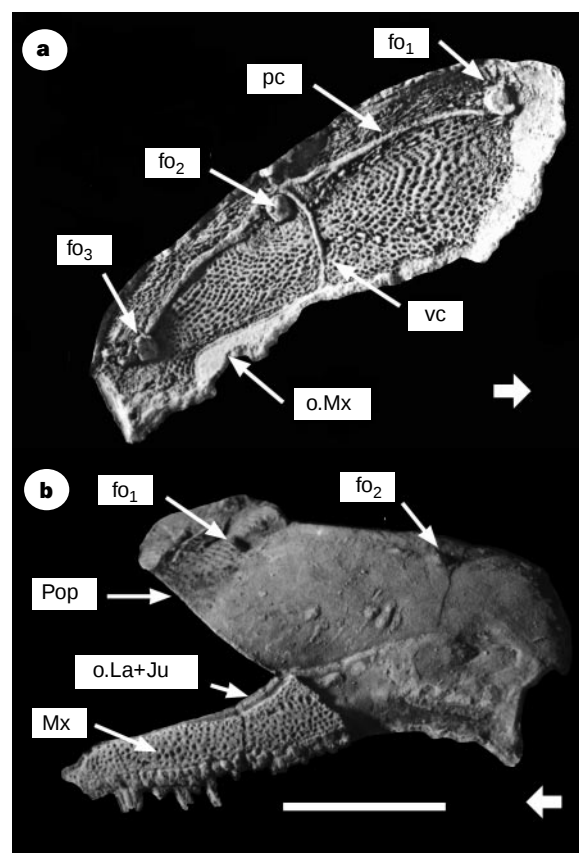


Figure 3 Cheek bones of *Psarolepis*. **a**, Specimen IVPP V11256.2: external mould of a left preopercular with the internal cast of sensory canals (pc, vc) and the pore-canal system, from the third bed in the Xishancun Formation. **b**, Specimen IVPP V11255: lateral view of a left preopercular (Pop) and maxillary (Mx), from the second bed in the Xishancun Formation. Short, thick arrows point to the anterior end of the fish. Scale bar, 10 mm. See Fig. 1e for a simplified reconstruction. fo₁–fo₃, foramina of unknown function; o-La + Ju, area overlapped by lacrimal and jugal; o-Mx, area overlapped by maxillary; pc, internal cast of preopercular canal; vc, internal cast of vertical canal.

ref. 14. The strict consensus tree based on the modified codings in ref. 4 places *Psarolepis* as the sister group of all osteichthyans, whereas the strict consensus tree based on the original codings in ref. 14 places *Psarolepis* as the sister group of all previously known sarcopterygians (Fig. 4a, b). Figure 4c shows the two possible positions of *Psarolepis* and the incongruous distribution of *Psarolepis* features among the major gnathostome groups.

Although the clades common to the two competing schemes^{4,14} of sarcopterygian interrelationships remain well supported, the conflicts between the two schemes remain unresolved and the exact position of *Psarolepis* remains uncertain. The uncertainty results partly from a lack of information available for *Psarolepis* and other

important stem taxa in the data set, and partly from the difficulty of selecting and polarizing characters when both osteichthyan and non-osteichthyan groups are used in the same analysis. However, whether *Psarolepis* turns out to be a stem-group osteichthyan or a stem-group sarcopterygian, its unique character combination will have a marked impact on present studies of osteichthyan evolution. For instance, porolepiform-like features found in *Psarolepis* (a lower jaw with three infradentary foramina, well developed internasal cavities and parasymphysial areas carrying tooth whorls³) can no longer be used to define porolepiforms¹⁶ and/or dipnomorphs¹⁴ (porolepiforms and lungfishes). The polyplocodont folded teeth and the quadrostian skull roof pattern of osteolepiforms^{3,17} should

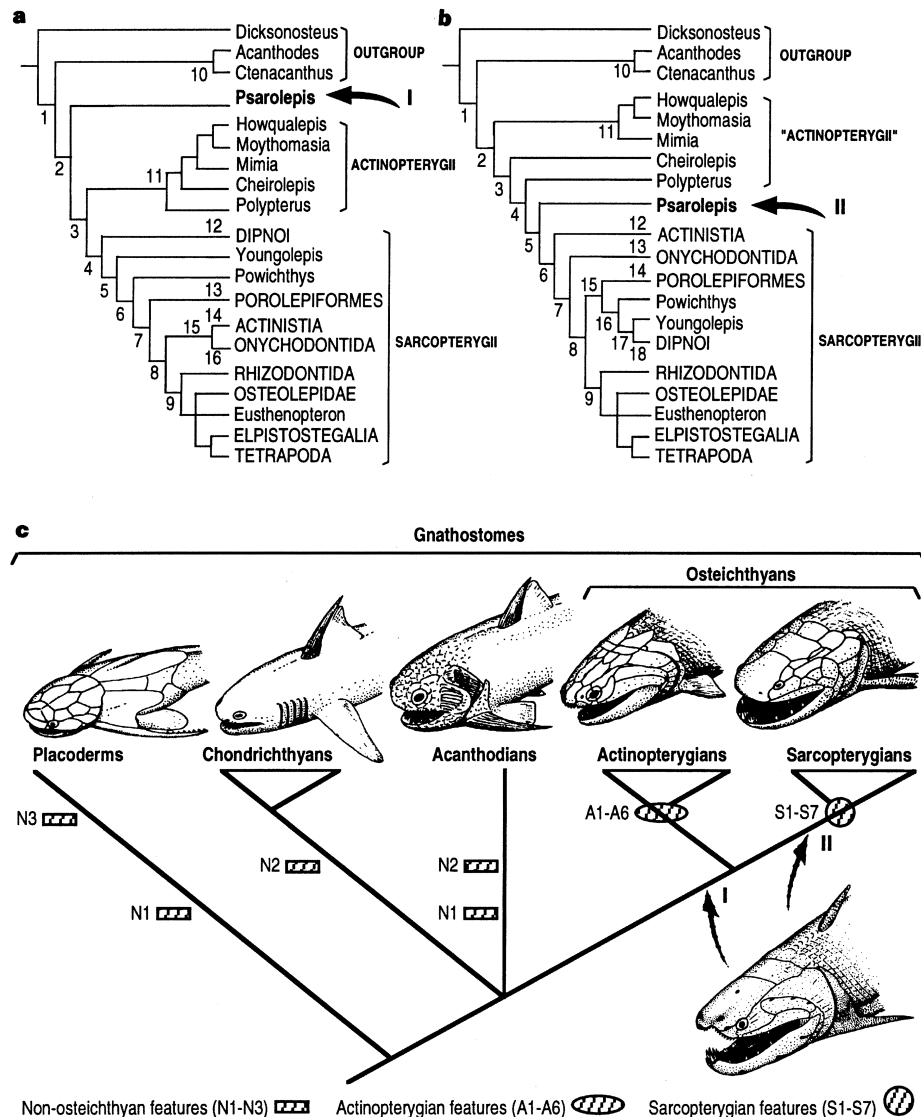


Figure 4 Phylogenetic analysis and the incongruous distribution of *Psarolepis* characters. **a**, Strict consensus tree from the expanded character matrix based on data from ref. 4 (54 most parsimonious trees at 311 steps, consistency index = 0.543, retention index = 0.804). *Psarolepis* forms the sister group of all osteichthyans. **b**, Strict consensus tree from the expanded character matrix based on data from ref. 14 (54 most parsimonious trees at 314 steps, consistency index = 0.538, retention index = 0.798). *Psarolepis* forms the sister group of all previously known sarcopterygians. The internal topologies of Sarcopterygii in **a**, **b** match the respective topologies in refs 4, 14 (see Methods for details). The trees are simplified by omitting lower-level nodes within the higher terminal taxa shown in uppercase letters. **c**, Interpreted inter-relationships of major gnathostome groups showing the incongruous distribution of *Psarolepis* characters. Our analysis uses limited non-osteichthyan characters to study the position of *Psarolepis* rather

than the inter-relationships of non-osteichthyan gnathostomes. The sister-group relationship between *Acanthodes* and *Ctenacanthus* in **a**, **b** is converted to a trichotomy between chondrichthyans, acanthodians and osteichthyans to reflect our interpretation. Features: N1, bony pectoral spines; N2, median fin spine; N3, endoskeletal shoulder girdle as massive plate pierced by openings for blood vessels and nerves; A1, toothed median rostral; A2, five coronoids in lower jaw; A3, absence of squamosal bone; A4, posteriorly expanded maxillary; A5, tilted preopercular with complete preopercular canal; A6, cleithrum with pointed vertical lamina; S1, lower jaw with three infradentary foramina; S2, well-developed internasal cavities; S3, parasymphysial area carrying tooth whorls; S4, polyplocodont teeth; S5, quadrostian skull roof pattern, S6, intracranial joint; S7, cosmine. See text for details.

also be regarded as primitive because of their presence in *Psarolepis*. If *Psarolepis* turns out to be a basal osteichthyan, the presence of an intracranial joint and cosmine can no longer serve as defining characters (synapomorphs) for sarcopterygians^{16–18}. □

Methods

Phylogenetic analysis. We added *Psarolepis* and three non-osteichthyan taxa (*Acanthodes*, an acanthodian, *Ctenacanthus*, a chondrichthyan, and *Dicksonosteus*, a placoderm) to the 33 taxa in ref. 14. We also added 9 characters to the 140 characters in ref. 14 to reflect the variations found in *Psarolepis* and the three non-osteichthyan outgroups (character 141, large dermal plates; 142, paired pectoral spines; 143, median fin spines; 144, denticulate postbranchial lamina of the cleithrum; 145, wide suborbital ledge; 146, eye stalk or unfinished area for similar structure; 147, ventral and otico-occipital fissures; 148, basiptyergoid articulation; 149, endochondral bone; 0 absent; 1 present). We adopted the same algorithm options as those used in refs 4, 14 (all characters unordered and unweighted) and used the three non-osteichthyan outgroups to root the trees. See Supplementary Information for the expanded matrix and for characters supporting major nodes in Fig. 4a, b. See ref. 14 for the original 140 characters and character states; see ref. 4 for the changed codings for characters 10, 17, 78 and 108.

Sarcopterygii, Dipnoi, Porolepiformes, Actinistia, Onychodontida and Tetrapodomorpha (including Rhizodontida, 'Osteolepiformes', Elpistostegalia and Tetrapoda) remain well supported in both trees. However, *Psarolepis* has changed the distribution and significance of many characters used previously to define osteichthyan groups (see Supplementary Information). In Fig. 4a, Sarcopterygii is defined by ten synapomorphies instead of fourteen as in ref. 14. Osteichthyes has no synapomorphy and is supported only by homoplasies. Actinopterygii is defined by one synapomorphy (character 6) and three reversals (characters 52, 93 and 110). In Fig. 4b, Sarcopterygii is defined by four synapomorphies, and 'Actinopterygii' appears as a paraphyletic group. One synapomorphy (character 7) defines the clade *Minimia* + (*Howqualepis* + *Moythomasia*) and five synapomorphies (characters 46, 63, 69, 98 and 134) define the clade *Polypterus* + (*Psarolepis* + Sarcopterygii). The position of *Polypterus* calls for further study, and the impact of data sampling and character coding on osteichthyan phylogeny deserves more attention.

Received 29 September 1998; accepted 7 January 1999.

- Janvier, P. *Early Vertebrates* (Oxford Univ. Press, Oxford, 1996).
- Rosen, D. E., Forey, P. L., Gardiner, B. G. & Patterson, C. Lungfishes, tetrapods, paleontology and plesiomorphy. *Bull. Am. Mus. Nat. Hist.* **167**, 159–276 (1981).
- Yu, X. A new porolepiform-like fish, *Psarolepis romeri*, gen. et. sp. nov. (Sarcopterygii, Osteichthyes) from the Lower Devonian of Yunnan, China. *J. Vert. Paleontol.* **18**, 261–274 (1998).
- Zhu, M. & Schultz, H.-P. The oldest sarcopterygian fish. *Lethaia* **30**, 293–304 (1997).
- Tong-Duzy, T., Ta-Hoa, P., Boucot, A. J., Goujet, D. & Janvier, P. Vertébrés siluriens du Viet-nam central. *C.R. Acad. Sci. Abstr.* **32**, 1023–1030 (1997).
- Gardiner, B. G. The relationships of the palaeoniscid fishes, a review based on new specimens of *Mimia* and *Moythomasia* from the Upper Devonian of Western Australia. *Bull. Brit. Mus. Nat. Hist.* **37**, 173–428 (1984).
- Otto, M. Zur systematischen Stellung der Lophosteiden (Obersilur, Pisces inc. sedis). *Paläont. Z.* **65**, 345–350 (1993).
- Schultze, H.-P. Ausgangform und Entwicklung der rhombischen Schuppen der Osteichthyes (Pisces). *Paläont. Z.* **51**, 152–168 (1977).
- Janvier, P. On the oldest known teleostome fish *Andreolepis hedei* Gross (Ludlow of Gotland), and the systematic position of the lophosteids. *Eesti NSV Teaduste Akadeemia Toimetised Geol.* **27**, 86–95 (1978).
- Chang, M. M. in *Early Vertebrates and Related Problems of Evolutionary Biology* (eds Chang, M. M., Liu, Y. H. & Zhang, G. R.) 355–378 (Science Press, Beijing, 1991).
- Chang, M. M. & Zhu, M. A new Middle Devonian osteolepidid from Qujing, Yunnan. *Mem. Ass. Australas. Palaeontol.* **15**, 183–198 (1993).
- Jessen, H. L. Lower Devonian Porolepiformes from the Canadian Arctic with special reference to *Powichthys thorsteinsoni* Jessen. *Palaeontogr. Abt. A* **167**, 180–214 (1980).
- Ahlberg, P. E. A re-examination of sarcopterygian interrelationships, with special reference to the Porolepiformes. *Zool. J. Linn. Soc.* **103**, 241–287 (1991).
- Cloutier, R. & Ahlberg, P. E. in *Interrelationships of Fishes* (eds Stiassny, M. L., Parenti, L. R. & Johnson, G. D.) 445–480 (Academic, New York, 1996).
- Swofford, D. L. PAUP: phylogenetic analysis using parsimony, version 3.1.1 (1993).
- Jarvik, E. Basic structure and evolution of vertebrates. Vol. 2 (Academic, London, 1981).
- Andrews, S. M. in *Interrelationships of Fishes* (eds Miles, R. S. & Patterson, C.) 137–177 (Academic, London, 1973).
- Bjerring, H. The 'intercranial joint' vs the 'ventral otic fissure'. *Act. Zool. Stockh.* **59**, 203–214 (1978).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank H.-P. Schultze, M. M. Chang and P. E. Ahlberg for useful discussions; W. Harre for the photographs and P. E. Ahlberg, M. I. Coates and M. M. Smith for comments and suggestions. M.Z. acknowledges support from the Alexander von Humboldt Foundation and the Chinese academy of Sciences. X.Y. thanks IVPP for access to specimens and Kean University for support for research and faculty development.

Correspondence and requests for materials should be addressed to M.Z. (e-mail: zhumin@ht.rol.cn.net).

Influence of motion signals on the perceived position of spatial pattern

Shin'ya Nishida* & Alan Johnston†

* Human and Information Science Research Laboratory, NTT Communication Science Laboratories, 3-1, Morinosato-Wakamiya, Atsugi-shi, Kanagawa 243-0198, Japan

† Department of Psychology and Institute of Cognitive Neuroscience, University College London, Gower Street, London WC1E 6BT, UK

After adaptation of the visual system to motion of a pattern in a particular direction, a static pattern appears to move in the opposite direction—the motion aftereffect (MAE)^{1,2}. It is thought that the MAE is not accompanied by a shift in perceived spatial position of the pattern being viewed^{3,4}, providing psychophysical evidence for a dissociation of the neural processing of motion and position that complements anatomical and physiological evidence of functional specialization in primate and human visual cortex^{5–7}. However, here we measure the perceived orientation of a static windmill pattern after adaptation to rotary motion and find a gradual shift in orientation in the direction of the illusory rotation, though at a rate much lower than the apparent rotation speed. The orientation shift, which started to decline within a few seconds, could persist longer than the MAE, and disappeared when the MAE was nulled by physical motion of the windmill pattern. Our results indicate that the representation of the position of spatial pattern is dynamically updated by neurons involved in the analysis of motion.

In the first experiment (see Methods), we investigated whether the MAE is accompanied by corresponding changes in apparent spatial position, and, if so, in what way the position shift changes over time. The results (Fig. 1) showed that after adaptation to a rotating windmill, a test windmill appeared to have rotated in the direction of the MAE. Using vernier tasks, T. Takeuchi (personal communication) and Snowden⁸ have found shifts in position after motion adaptation. However, we also found that the shift in perceived orientation, which presumably resulted from local changes in apparent position, gradually increased over the first few seconds. An incremental change began at the test onset rather than at the adaptation offset, a finding reminiscent of the storage effect of the MAE⁹. Over the first two seconds, the orientation shift increased almost linearly; thus, for this interval, we can directly compare the rate of orientation change with the MAE speed. The gain, that is, the slope of the functions in Fig. 1 (0.40 and 0.65 degrees s^{−1} for subjects SN and AJ, respectively) divided by the MAE speed measured separately (5.05 and 8.00 degrees s^{−1}, respectively), was ~8%. These results show that the shift in orientation is not simply an alternative measure of the MAE and are consistent with the classical view that the perception of motion and that of position change are dissociated³. We rejected the idea that the magnitude of the orientation shift might be proportional to MAE strength, because we found that MAE speed did not increase as a function of time over the same interval (data not shown).

The shift in orientation started to decline a few seconds after the test onset. We measured the shift for a stationary test at repeated intervals for up to 2 min, and compared the decay function with that of the MAE (Fig. 2a) (experiment 2; see Methods). For two subjects the orientation shift decayed more slowly and persisted for longer than the MAE. Significant amounts of orientation shift remained even after the MAE finished. However, for another naïve subject, the decay functions of the orientation shift and the MAE were quite similar.

We attribute the orientation shift to an adjustment in the representation of the position of spatial pattern by motion information. Although the adjustment is driven by the motion system, it does not occur through direct linkage to the motion system, because the shift increases at first whereas the MAE always declines, and the shift can last longer than the MAE. We explain this dissociation by suggesting that the motion system induces incremental orientation shifts that are integrated over time within the system that represents spatial pattern, although motion signals can induce a position shift without temporal integration, as the test windmill appeared to be rotated even at test onset. The shift at test onset may be identified with direct effects of motion signals on spatial position. The small slope of the initial increase as well as the subsequent gradual decay can be attributed to the effects of pattern input, which acts to restore the veridical correspondence between the spatial representation and its spatial input. This interpretation implies that the representation of spatial pattern involves a dynamic system with a finite memory that integrates information over time, rather than being simply a reflection of the current spatial input at the eye.

This argument led us to expect that an established spatial representation may be reset in the face of a radical change of the spatial input (for example, presentation of a new stimulus). To test this we repeated experiment 2 (see Methods), this time reversing the phase of the test windmill in the intervals between orientation judgements. If the spatial representation is refreshed at phase

reversals, properties arising from temporal integration should be selectively disrupted whereas direct effects of motion signals on spatial position should remain unchanged. As expected, we found that the orientation shift now declined in step with the MAE for all subjects (Fig. 2b). The phase reversal had no effect for the naïve subject whose position shift decayed rapidly in the initial experiment, which suggests that this subject occasionally refreshed his representation without any change in the test stimulus, perhaps by eye movements or attentional control.

Several studies^{10–12} have shown that moving stimuli appear to be spatially displaced in the direction of motion, but the underlying mechanism of this phenomenon remains controversial^{13–15}. We measured relative mean orientation for briefly presented pairs of test stimuli that rotated in opposite directions at a range of velocities (experiment 3; see Methods). Perceived orientation was shifted in the direction of rotation by an amount that was roughly proportional to the velocity (Fig. 3). After adaptation, orientation judgements were biased in the direction opposite to that of the adapting stimulus, as before. However, when the test rotation speed was such that it nulled the MAE, the spatial shift disappeared. Thus a moving pattern does not appear to be spatially displaced if it appears static, which suggests a common source for mislocalization effects induced by physical motion and motion adaptation.

The fact that spatial shifts do occur for physically static but apparently moving test patterns seems to be incompatible with an

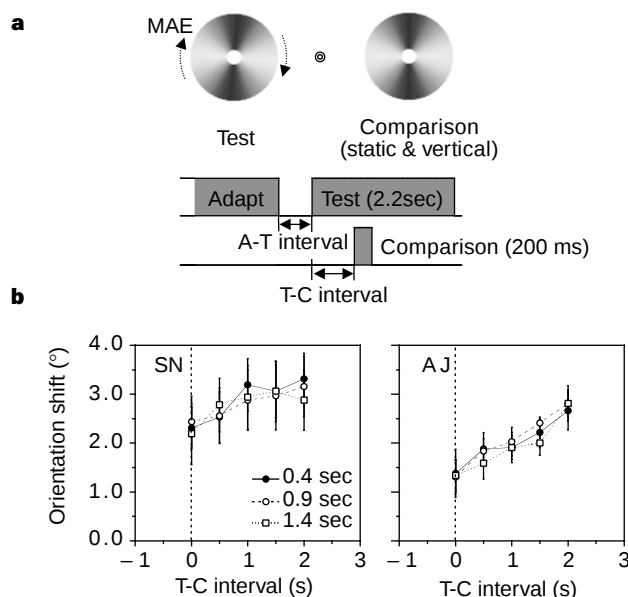


Figure 1 Experiment 1. **a**, The experimental procedure. Top, after adaptation of the visual system to a rotating windmill (not shown), a static test windmill (left) and a comparison windmill (right) on the opposite side of fixation (centre) were presented. The test windmill appeared to have rotated in the direction of the MAE, and the magnitude of the shift gradually increased over time. Bottom, the A-T interval (time between adaptation offset and test onset, and the T-C interval (time between test onset and onset of the comparison stimulus). **b**, Results for subjects SN and AJ. The ordinate shows the amount of rotation in the direction opposite to the MAE required to make the test stimulus appear parallel to the comparison stimulus. The abscissa shows the T-C interval. The symbols indicate variation in the A-T interval. Each point is based on eight measurements (error bar: ± 1 s.e.m.). Similar results were obtained with another naïve subject (PS).

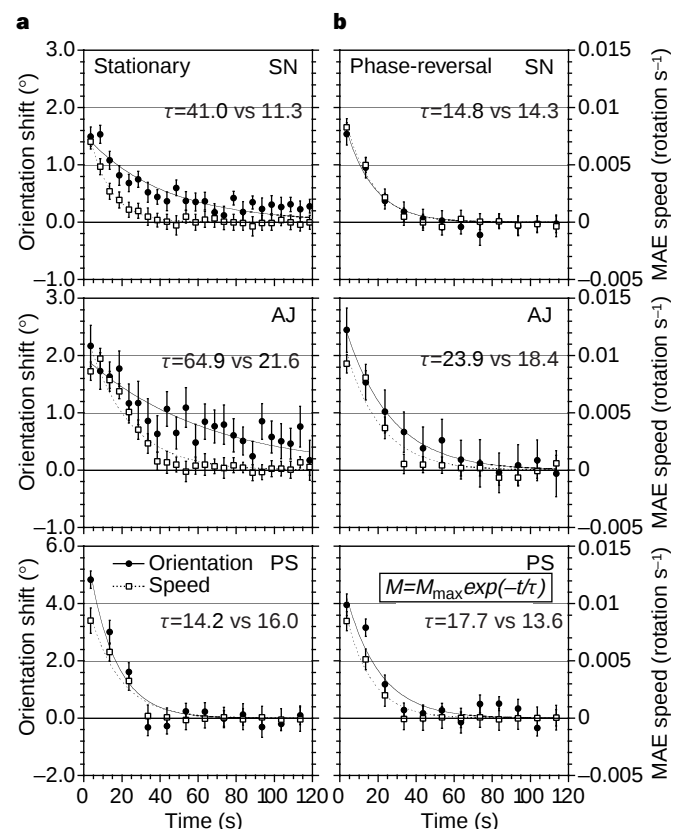


Figure 2 Experiment 2. Comparison of the decay function of the orientation shift to that of the MAE when the same test stimulus was continuously presented and when the phase of the test was reversed during interjudgement intervals. **a**, Continuous presentation of test symbols; **b**, reversal of the test phase during interjudgement intervals. The results show that the orientation shift (circles) could persist longer than the MAE (squares) in the former condition (**a**), but not in the latter (**b**). Each point is based on at least eight measurements (error bar: ± 1 s.e.m.). The smooth curve is the best fitting exponential function, $M = M_{\max} \exp(-t/\tau)$. The time constant (τ) of each function is shown; the first number refers to the orientation shift and the second to the MAE.

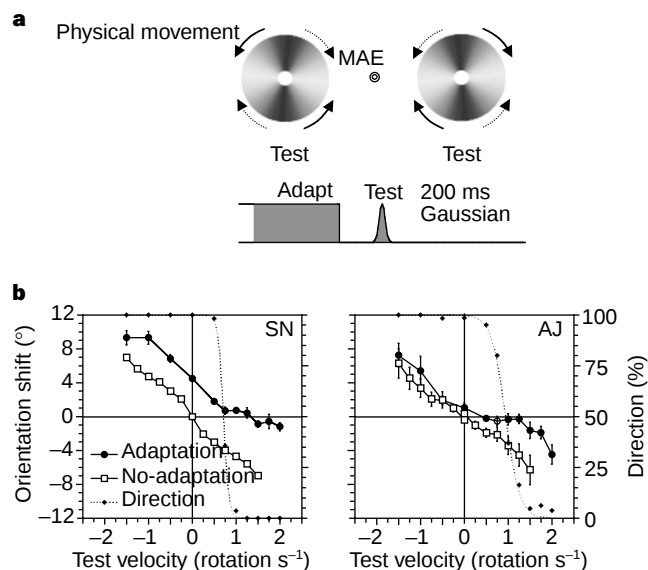


Figure 3 Experiment 3. **a**, The experimental procedure. After adaptation to windmills presented on both sides of fixation, two test windmills, rotating at a given speed, were presented within a 200-ms gaussian window. **b**, the orientation shift of the physically rotating windmill as a function of the rotation speed before (squares) and after (circles) motion adaptation is plotted alongside a psychometric function showing judgements of direction measured after adaptation (diamonds). Without adaptation, the average test orientation appeared to be shifted in the direction of real movement (downwards for positive test speeds). After adaptation, perceived orientation was shifted in the aftereffect direction (upward), and the absolute orientation shift almost disappeared when the MAE was nulled by real motion (that is, the test appeared stationary). Each point is based on at least eight measurements. Similar results were obtained with subject PS.

explanation of the mislocalization effect in terms of a possible difference in peripheral-processing time delay between moving and static stimuli^{13,14}. Another explanation for mislocalization, that motion sensors may systematically misreport their position towards the sensor's preferred direction¹⁵, incorrectly predicts that the magnitude of mislocalization should always vary with the activation of the motion sensors (MAE magnitude), whereas we found several examples of dissociation of the magnitude of mislocalization and the MAE magnitude. Finally, motion and position may interact at various levels in the visual system^{16,17}. Given that the most precise representation of local spatial information exists in early cortical areas, an intriguing possibility is that our findings may reflect recurrent input from cortical areas V5/MT to V1 (ref. 18) or V2 (ref. 19). □

Methods

Stimuli were presented on a ManiTron monitor (P31 phosphor, 60 Hz refresh) under control of an IBM compatible PC with a VSG2/3 board. Viewing was binocular from 93 cm. Windmills subtended 4° in diameter with a 40-min hole at the centre and were presented 4° left or right of a fixation bullseye. Each windmill had a sinusoidal luminance modulation of 2 cycles per rotation at 80% contrast on a 6.5 cd m⁻² background.

Experiment 1. After a 3-min adaptation to a windmill moving at 3 rotations s⁻¹ (6 Hz), a static test windmill was presented for 2.2 s at the same location, and a comparison windmill was presented for 0.2 s on the opposite side of the fixation, both with sharp onset and offset. Subsequent test presentations were always preceded by an 8-s re-adaptation period. Subjects judged whether the dark region of the test windmill was rotated in the clockwise or anti-clockwise direction relative to a comparison windmill in which the dark region was always vertical. The use of a relative judgement allowed us to discount any influence of eye movements. The shift was estimated using a standard staircase program that adjusted the test orientation until at least four reversals of direction had

occurred. The time of onset of the comparison stimulus (T–C interval) and the time between offset of the adaptation and onset of the test (A–T interval) was varied between trials. To equate the state of adaptation, the staircase sequences, for each stimulus condition, were intermingled within a single block, with each staircase waiting at the *n*th reversal point until all the other staircases reached that point. The adaptation direction and position were changed between blocks.

Experiment 2. In the first experiment 2, the temporal decay of the magnitude of the position shift was measured by the method of adjustment. The test stimulus was presented for 2 min after 1 min of adaptation to a rotating windmill and a 1-s blank interval. During the test presentation, a comparison windmill was presented on the opposite side of fixation for 4 s every 10 s. The subject had to adjust the orientation of the comparison to match it with the test. The perceived orientation was estimated from the settings during the last 1 s of every 4-s matching period. The test orientation as well as the initial comparison orientation for each matching period were randomly jittered within ±2.5° of the vertical. There was at least a 1-min interval between blocks. The direction and position of the adaptation stimulus were changed between blocks. To halve the sampling interval, the first matching period started at the onset of the test in half of the blocks, and 5 s later in the other half. When repeating experiment 2, we used the same procedure as before, except that, although the test stimulus was always vertical (with a small jitter) during the matching period, its orientation switched to the horizontal 0.5 s after a match was recorded, and then returned to the vertical 0.5 s before the next matching period. The first matching period always started at test onset.

Experiment 3. After adaptation to windmills presented on both sides of fixation, two test windmills, rotating at a given speed, were presented within a 200-ms gaussian temporal window (s.d. = 35.4 ms, peaking at 1.5 s after the adaptation offset). The two windmills were always mirror-image symmetric with respect to a virtual central vertical line. Subjects compared the average orientation of the two test stimuli. We used a staircase method to estimate the perceptually parallel angle, adjusting orientation according to the subject's response. To find the motion null point, we also asked the subject to judge the direction of perceived motion. If, on occasion, the test stimulus appeared to move first in the MAE direction and then in the other direction, subjects were instructed to choose the dominant direction. In each block, staircases for each test velocity were intermingled. A similar procedure was used for the no-adaptation condition, except that a uniform field was presented for 3 s between trials and the judgement of rotation direction was not required.

Received 9 September; accepted 12 December 1998.

1. Wohlgenuth, A. On the aftereffect of seen movement. *Br. J. Psychol. Monogr. Suppl.* **1**, 1–117 (1911).
2. Mather, G., Verstraten, F. A. J. & Anstis, S. M. *The Motion Aftereffect: A Modern Perspective* (MIT Press, Cambridge, Massachusetts, 1998).
3. Gregory, R. L. *Eye and Brain* (McGraw-Hill, New York, 1966).
4. Nakayama, K. Biological image motion processing: a review. *Vision Res.* **25**, 625–660 (1985).
5. Zeki, S. M. Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *J. Physiol. (Lond.)* **236**, 549–573 (1974).
6. Zihl, J., VonCramon, D. & Mai, N. Selective disturbance of movement vision after bilateral brain damage. *Brain* **106**, 313–340 (1983).
7. Livingstone, M. S. & Hubel, D. H. Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *J. Neurosci.* **7**, 3416–3466 (1987).
8. Snowden, R. J. Perceived spatial position is altered by motion adaptation. *Perception* **27** (Suppl.), 49–50 (1998).
9. Spigel, I. M. The effects of differential post exposure illumination on the decay of a movement aftereffect. *J. Psychol.* **50**, 209–210 (1960).
10. Ramachandran, V. S. & Anstis, S. M. Illusory displacement of equiluminous kinetic edges. *Perception* **19**, 611–616 (1990).
11. De Valois, R. L. & De Valois, K. K. Vernier acuity with stationary moving Gabors. *Vision Res.* **31**, 1619–1626 (1991).
12. Nijhawan, R. Motion extrapolation in catching. *Nature* **370**, 256–267 (1994).
13. Baldo, M. V. C. & Klein, S. A. Extrapolation or attentional shift. *Nature* **378**, 565–566 (1995).
14. Khurana, B. & Nijhawan, R. Extrapolation or attentional shift: reply. *Nature* **378**, 566 (1995).
15. Anstis, S. M. & Remachandran, V. S. In *The Artful Eye* (eds Gregory, R., Harris, J., Heard, P. & Rose, D.) 232–248 (Oxford Univ. Press, Oxford, 1995).
16. Merigan, W. H. & Maunsell, J. H. R. How parallel are the primate visual pathways? *Annu. Rev. Neurosci.* **16**, 369–402 (1993).
17. Zeki, S. *A Vision of the Brain* (Blackwell, Oxford, 1993).
18. Shipp, S. & Zeki, S. The organisation of connections between areas V1 and V5 in the macaque monkey visual cortex. *Eur. J. Neurosci.* **1**, 309–332 (1989).
19. Shipp, S. & Zeki, S. The organisation of connections between areas V2 and V5 in the macaque monkey visual cortex. *Eur. J. Neurosci.* **1**, 333–354 (1989).

Acknowledgements. S.N. was a visiting fellow at the Institute of Cognitive Neuroscience, UCL. We thank S. Naito and K. Ishii, NTT Communication Science Laboratories for their support. A.J. is supported by a grant from the EPSRC/BBSRC Mathematical Modelling Initiative.

Correspondence and requests for materials should be addressed to A.J. (e-mail: a.johnston@ucl.ac.uk).

The global record of memory in hippocampal neuronal activity

Emma R. Wood, Paul A. Dudchenko & Howard Eichenbaum

Laboratory of Cognitive Neurobiology, Department of Psychology, Boston University, 64 Cummington Street, Boston, Massachusetts 02215, USA

In humans the hippocampal region of the brain is crucial for declarative¹ or episodic² memory for a broad range of materials. In contrast, there has been controversy over whether the hippocampus mediates a similarly general memory function in other species, or whether it is dedicated to spatial memory processing^{3–6}. Evidence for the spatial view is derived principally from the observations of 'place cells'—hippocampal neurons that fire whenever the animal is in a particular location in its environment^{7,8}, or when it perceives a specific stimulus or performs a specific behaviour in a particular place^{3,4}. We trained rats to perform the same recognition memory task in several distinct locations in a rich spatial environment and found that the activity of many hippocampal neurons was related consistently to perceptual, behavioural or cognitive events, regardless of the location where these events occurred. These results indicate that non-spatial events are fundamental elements of hippocampal representation, and support the view that, across species, the hippocampus has a broad role in information processing associated with memory.

Rats were trained to perform an odour-guided, continuous, non-matching-to-sample task at nine different locations on an open platform (Fig. 1). On each trial a cup containing sand scented with one of nine spices (for example, thyme, paprika) was placed on the platform. On half of the trials the scent was different from that of the previous trial (a 'non-match') and a cereal reward was buried at the bottom of the cup. On the remaining trials the scent was repeated (a 'match') and the cup was not baited. Rats approached each cup, sniffed the sand, and then either dug for the reward or turned away. After either response, the cup was removed and there was a 10-s interval before the next trial, during which the rat remained on the platform. The location of the cup changed on each trial and did not predict whether the trial was a non-match or a match. The cup

Table 1 Cells with significant task-related correlates

Nonspatial variables	51	Spatial variables	40
Approach	26	Position	14
Odour	10	Position and Odour	4
M/NM*	13	Position and M/NM*	18
Odour and M/NM*	2	Position and Odour and M/NM*	4

The activity of 91 cells (out of 127) was statistically associated with the variables tested.

* M/NM, match/non-match trial type.

location, sequence of odours and match/non-match contingencies followed a pseudo-random order, such that each session comprised 108 trials, with 12 presentations of each odour and each cup location. This design enabled us to dissociate hippocampal neural activity associated with approaching the stimulus cups, the different odours, and the match and non-match trial types from the locations where these events occurred.

After being trained to dig in sand, rats learned within three testing sessions to perform the differential response that involved either digging through the sand to find the buried reward on non-match trials, or turning away without digging on the match trials (a range of 100–280 trials was required to perform correct responses on 90% of trials over 20 consecutive trials). They continued to perform the task consistently during the recording sessions (mean 96.8% correct (range 88–100%) for each 108-trial session). On all of several non-match probe trials in which the reward was omitted, each rat dug appropriately, confirming that their accurate performance was not based on detecting the food directly.

The activity of 127 single units identified as complex spike cells⁹ was recorded from the CA1 and CA3 pyramidal cell fields of four rats performing the non-matching-to-sample task. The activity of 91 cells was statistically associated with one or more of the variables tested, and over half of these were associated only with non-spatial variables (Table 1). We identified ten cells that fired differentially across the different odour stimuli, but not across the match/non-match trial types or across the different locations where the trials were performed. For example, the cell depicted in Fig. 2a responded to odour 5 more than to any other odour, and was not influenced by trial type or cup location. To examine the robustness of the discrimination among odours for all cells of this category, we normalized their firing rates and compared average responses to the best two odours (those associated with the highest firing rates)

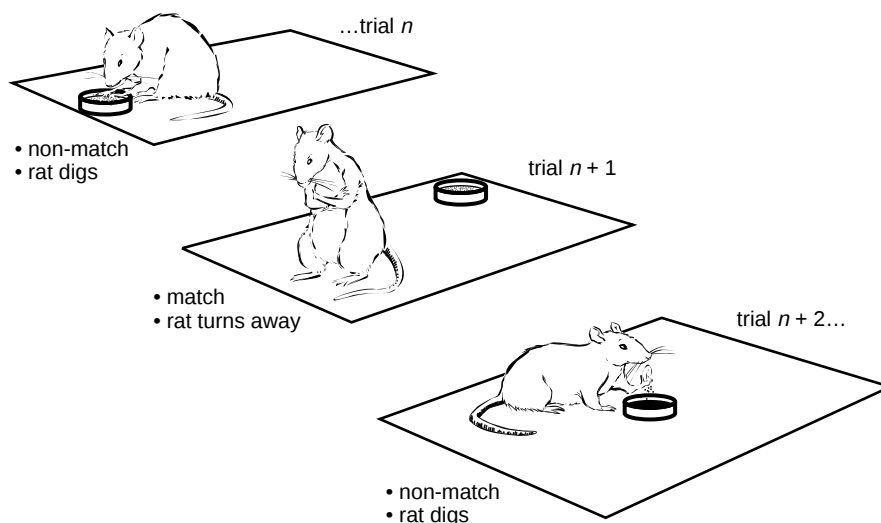


Figure 1 Continuous non-matching to sample task. Trial n represents a non-match trial where the odour differs from that presented on the previous trial, and the rat digs to find a buried reward. On the next trial ($n + 1$), the same scent is repeated, although in a different location. As no reward is available, animals learn

not to dig on these match trials and to turn away from the cup. On the subsequent trial ($n + 2$), the odour again differs from that of the previous trial and the animal digs for a buried reward. Note that the position of the cup is independent of the match/non-match contingency.

with responses to the worst two odours (lowest firing rates). Odour cells strongly differentiated best and worst odours (Fig. 3a, left panel). To examine whether odour-selective activity was present across the different locations, we compared the normalized responses among positions rank-ordered for each cell according to the magnitude of the overall firing rate associated with each position (see Methods). Although the magnitude of the odour specificity declined across the rank-ordered positions, odour cells discriminated best and worst odours at all positions (Fig. 3a, right panel).

We also identified 14 cells that fired differentially across the different positions of the trial, but not across the different odours or match/non-match trial types. For example, the cell depicted in Fig. 2b fired at higher rates when the rat performed trials at positions P2 and P3 than at other positions, but was not influenced

by the trial type or odour. As a group, location cells strongly discriminated their two best positions (those associated with their highest firing rates) from their two worst positions (Fig. 3b, left panel). Furthermore, although the magnitude of location specificity declined across the rank-ordered odour series, the ability of location cells to discriminate positions was found across the whole odour set (Fig. 3b, right). Notably, the magnitude of the normalized firing rates and of the discrimination between best and worst conditions was similar for odour cells and location cells (compare panels in Fig. 3a, b), indicating that the robustness of odour responses is as great as that for location responses within these analyses.

In addition, 13 cells fired differentially between match and non-match trials, but not across the different odours or locations. For example, the cell depicted in Fig. 2c had a greater firing rate on match trials than on non-match trials, regardless of the odour or

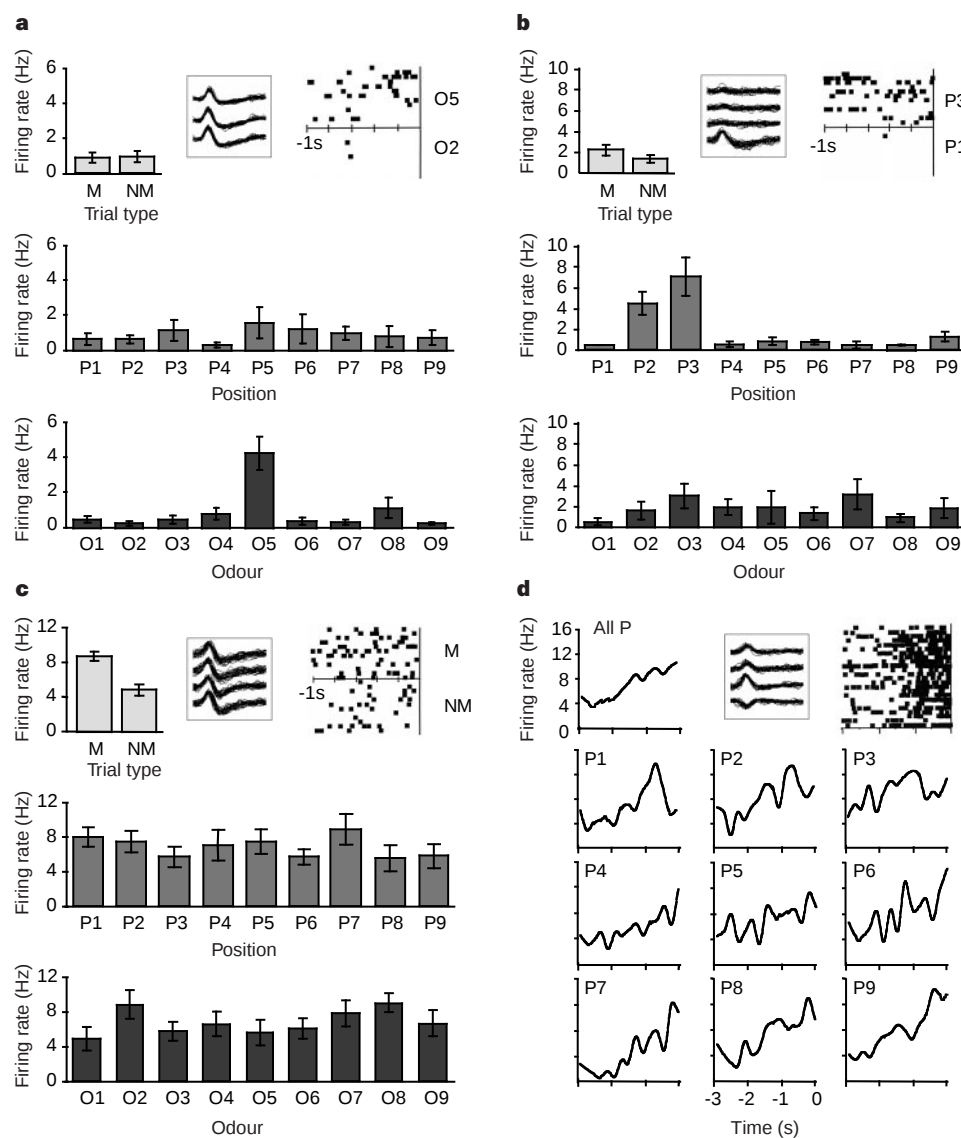


Figure 2 Examples of task-related firing patterns of hippocampal neurons. **a–c**, Firing rate in 1-s analysis period for each trial type (M, match; NM, non-match), cup location (P1–P9) and odour (O1–O9) for: **a**, an odour cell (odour, $F_{(8,74)} = 8.59$, $P < 0.0001$; trial type, $F_{(1,74)} = 0.04$, not significant (NS); cup location $F_{(8,74)} = 1.03$, NS; odour $\times$ trial type, $F_{(8,74)} = 0.75$, NS; location $\times$ trial type, $F_{(8,74)} = 1.74$, NS); **b**, a location cell (cup location, $F_{(8,74)} = 8.60$; $P < 0.0001$; odour, $F_{(8,74)} = 0.84$, NS; trial type $F_{(1,74)} = 2.76$, NS; odour $\times$ trial type, $F_{(8,74)} = 1.14$, NS; cup location $\times$ trial type, $F_{(8,74)} = 1.58$, NS); and **c**, a match cell (trial type, $F_{(1,74)} = 22.95$, $P < 0.0001$; odour, $F_{(8,74)} = 1.42$, NS; location, $F_{(8,74)} = 1.17$, NS; odour $\times$ trial type, $F_{(8,74)} = 0.68$,

NS; location $\times$ trial type, $F_{(8,74)} = 1.20$, NS). **d**, Firing rates (200-ms bins) for 3-s period when the rat approached each cup position (P1–P9), and averaged across all positions (All P) for an approach cell (trial period, $t_{(1,107)} = 10.77$, $P < 0.001$; trial type, $F_{(1,74)} = 0.06$, NS; odour $F_{(8,74)} = 0.47$, NS; location $F_{(8,74)} = 1.42$, NS; odour $\times$ trial type, $F_{(8,74)} = 0.96$, NS; location $\times$ trial type, $F_{(8,74)} = 1.00$, NS). Each panel also shows the waveform of the cell recorded on each tetrode channel and a raster display of firing patterns time-locked to the end of the odour sample period.

location where the trial was performed. Match/non-match cells as a group strongly discriminated the two trial types (Fig. 3c, left panel), and this effect was observed at all rank-ordered positions (Fig. 3c, right panel).

The firing rate of a further 26 cells changed significantly during the approach or arrival at the stimulus cup, but the cells did not fire differentially across the different odours, trial types, or trial locations. The cell depicted in Fig. 2d increased its firing rate as the rat arrived at each cup, regardless of any other variables. As a group, approach cells strongly differentiated the initiation of the trial and arrival at the stimulus cup (Fig. 3d, left panel), and this effect was observed at all rank-ordered positions (Fig. 3d, right panel).

Of the 127 cells analysed (Table 1), 51 (40.2%) had solely non-spatial correlates, which is reflected in a statistically significant main effect for odour, trial type or approach in the absence of a significant main effect or interaction involving cup position. Nearly all of these cells encoded just one of these non-spatial properties. A further 40 cells (31.5%) had a spatial correlate, reflected in a statistically significant main effect or interaction involving cup location. Only 14 of the spatial cells were pure location cells. The remaining 26 spatial cells, as well as two non-spatial cells, encoded combinations of multiple variables. The activity of the remaining 36 cells (28.3%) was not significantly affected by any of the variables measured. This overall pattern of results indicates that non-spatial and spatial

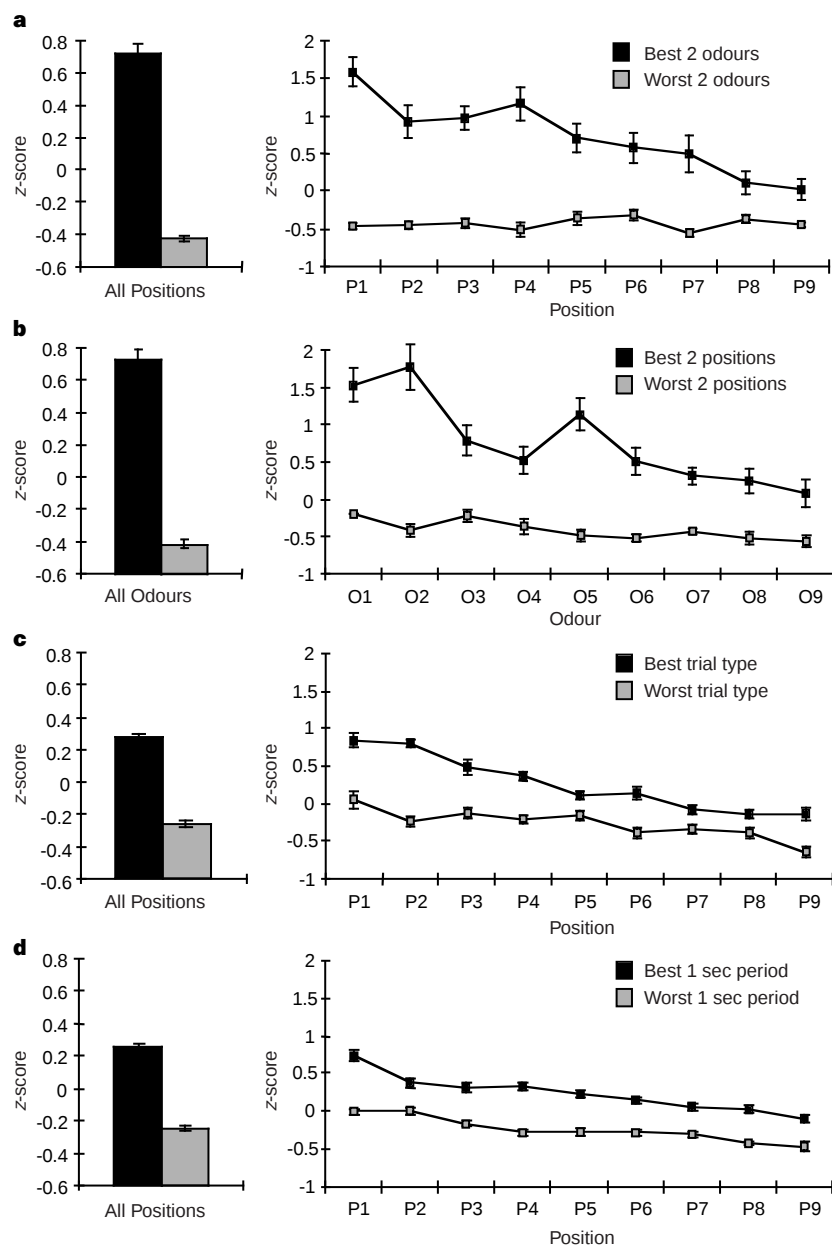


Figure 3 Different groups of hippocampal neurons discriminate odours, positions, match/non-match trial types and approach period. **a**, Odour cells ($n = 10$); **b**, location cells ($n = 14$); **c**, Match/non-match cells ($n = 13$); **d**, approach cells ($n = 26$). Left: comparison of mean ($\pm$ s.e.) normalized firing rates to **a**, the best and worst two odours ($t_{(1,9)} = 19.75$, $P < 0.0001$); **b**, the best and worst two positions ($t_{(1,13)} = 14.37$, $P < 0.0001$); **c**, the best and worst trial type ($t_{(1,12)} = 14.34$, $P < 0.0001$); and **d**, the best and worst approach period ($t_{(1,25)} = 13.89$, $P < 0.0001$) for all cells in that category. Right: normalized responses compared across the rank-ordered positions (P1–P9; **a**, **c**, **d**) or rank-ordered odours (O1–O9; **b**). **a**, odour,

$F_{(1,72)} = 527.43$, $P < 0.0001$; position, $F_{(8,72)} = 3.92$, $P < 0.01$; odour $\times$ position, $F_{(8,72)} = 3.64$, $P < 0.05$; odour effect at each position, $P < 0.05$; **b**, position, $F_{(1,104)} = 231.57$, $P < 0.0001$; odour, $F_{(8,104)} = 7.67$, $P < 0.0001$; position $\times$ odour, $F_{(8,104)} = 3.91$, $P < 0.01$; position effect for each odour, $P < 0.05$; **c**, trial type, $F_{(1,96)} = 205.71$, $P < 0.0001$; position, $F_{(8,96)} = 139.44$, $P < 0.0001$; trial type $\times$ position, $F_{(8,96)} = 3.26$, $P < 0.05$; trial type effect at each position, $P < 0.05$; **d**, trial period, $F_{(1,200)} = 192.95$, $P < 0.0001$; position, $F_{(8,200)} = 136.88$, $P < 0.0001$; trial period $\times$ position, $F_{(8,144)} = 1.54$, NS; trial periods effect at each position, $P < 0.05$.

information are both fundamental components of hippocampal representations.

Previous studies have associated hippocampal neural activity with a variety of specific non-spatial stimuli and behaviours in animals performing various tasks^{10–20}. In most of these studies the events occurred at just one place, raising the possibility that the activity was location specific, or that the non-spatial events were defining features of the locations where they occurred^{10,11,13–19}. Other studies presented the same stimuli in multiple locations, but these stimuli were enclosures, landmarks or walls, which might have been perceived as distinctive spatial reference frames within the overall environment^{12,20–23}. In our study, punctate stimuli were transferred to and from the environment while the rat was present, emphasizing interactions with them as episodes occurring within a single, constant spatial reference frame. These results provide compelling evidence that hippocampal neurons represent information related to specific perceptual, cognitive and behavioural events not tied to a particular spatial location. Furthermore, the range of specificity varied from many cells that encode single, frequently experienced events, such as a place or an odour, to a few cells encoding unique conjunctions that reflect a rare event, such as a specific odour appearing as a 'match' in a particular place. Such a range of coding specificity could reflect the emergence of consistencies within a global knowledge structure from a record of specific episodes^{24,25}. □

Methods

Behavioural apparatus. Rats were tested on a square, black, plexiglass platform (92 cm × 92 cm) with 2-cm high walls along each side, elevated 70 cm above the floor. The stimuli were plastic cups (12 cm diameter and 4.5 cm deep) containing 100 g of sand scented with 1 g of a common spice such as cinnamon or cumin. The stimuli could be placed in any one of nine locations: in the four corners, half way along each side, or in the centre of the platform. The platform was located at one end of a large testing room, amid a rich array of constant distal spatial cues. See text for behavioural testing methods.

Electrophysiological recording. Electrophysiological recording techniques were as described previously²⁶. Briefly, four rats were surgically implanted with two or four moveable tetrodes just above the dorsal hippocampus (3.5 mm posterior and 2.3 mm lateral to bregma). Unit activity and the rat's location were recorded while rats were run on the odour-guided non-matching-to-sample task. Single units were isolated after the recording session by identifying clusters defined on the basis of waveform parameters (Autocut, Datawave Technologies). The recordings were considered stable if the clusters remained within the same fixed boundaries throughout the recording session (see example waveforms in Fig. 2). Cells with a signal-to-noise ratio of more than 3:1, a spike duration (peak time to valley time) of greater than 250 ms, a mean firing rate (total number of spikes divided by total session time) of less than 2.5 Hz, and a minimum spike count of 100 over the recording session were defined as complex spike cells. Electrode placements in CA1 and CA3 subfields were verified histologically.

Analysis of cellular activity. To assess whether each cell's activity was associated with odour, cup location or match/non-match trial type, we examined the last second of each trial when the rat arrived at the cup and sampled the scented sand, immediately before it either dug or turned away. Activity during this period was subjected to a three-way analysis of variance (ANOVA), which analysed the firing rates across the nine different odours, the nine cup locations and the two trial types (match and non-match), and the two-way interactions between odour and trial type, and between cup location and trial type. Cells with one significant main effect ($P < 0.05$) in the absence of any other main effects or interactions were categorized as odour, location or match/non-match cells accordingly. To assess changes in activity associated with approach to the cup, we compared (by using paired t -tests) the firing rate during the 1-s period beginning 3 s before the behavioural response, when the animal initiates its approach to the cup, with that during the 1 s immediately before the behavioural response. Cells with a significant ($P < 0.05$) t -value, and with no significant main effects or interactions in the three-way ANOVA, were classified as 'approach cells'.

To assess the robustness of coding for cells in each category (odour, location,

match/non-match and approach), we first normalized their responses by converting the firing rates during the critical 1-s periods for each trial to z -scores relative to the same periods for all trials. We then compared (using paired t -tests) the responses to the best with the worst relevant condition for each category of cells (best versus worst two odours or two positions, best versus worst match/non-match trial type, or best versus worst 1-s period during the approach). In addition, to determine whether odour, trial-type and approach cells had selective responses across all positions, we first rank-ordered the positions for each cell from that associated with the highest to the lowest firing rate during the 1-s trial period (averaged across 12 trials at each position). Then, for each cell type, we used a two-way ANOVA to compare the normalized responses to the best and worst conditions (as above) across the nine positions. This strategy necessarily results in an orderly decline in firing rates across the rank-ordered position series. Nevertheless, significant discrimination at all positions by odour, trial type or approach cells could be identified as a significant main effect and significant post-hoc effects for that variable at all positions (paired one-tailed t -tests). Using the same strategy, we also compared the normalized responses to the best and worst two positions for location cells across the nine rank-ordered odours. As with the other analyses, this necessarily results in an orderly decline of firing rates across the rank-ordered odour series. Nevertheless, this analysis allowed us to compare the consistency of the location responses to the responses for the other cell types.

Received 28 September; accepted 10 December 1998.

- Cohen, N. J. & Squire, L. R. Preserved learning and retention of pattern analyzing skill in amnesia: dissociation of knowing how and knowing that. *Science* **210**, 207–210 (1980).
- Vargha-Khadem, F. *et al.* Differential effects of early hippocampal pathology on episodic and semantic memory. *Science* **277**, 376–380 (1997).
- O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Oxford University Press, London, 1978).
- Nadel, L. The hippocampus and space revisited. *Hippocampus* **1**, 221–229 (1991) and accompanying debate, 230–291.
- Cohen, N. J. & Eichenbaum, H. *Memory, Amnesia and the Hippocampal System* (MIT Press, Cambridge, Massachusetts, 1993).
- Jarrard, L. E. On the role of the hippocampus in learning and memory in the rat. *Behav. Neural Biol.* **60**, 9–26 (1993).
- O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34**, 171–175 (1971).
- Muller, R. U., Kubie, J. L. & Ranck, J. B. Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *J. Neurosci.* **7**, 1935–1950 (1987).
- Ranck, J. B. Studies on single neurons in dorsal hippocampal formation and septum in unrestrained rats. Part 1. Behavioral correlates and firing repertoires. *Exp. Neurol.* **41**, 461–531 (1973).
- Berger, T. W., Rinaldi, P. C., Weisz, D. J. & Thompson, R. F. Single-unit analysis of different hippocampal cell types during classical conditioning of rabbit nictitating membrane response. *J. Neurophysiol.* **50**, 1197–1219 (1983).
- Weiss, C., Kronforst-Collins, M. A. & Disterhoft, J. F. Activity of hippocampal pyramidal neurons during trace eyeblink conditioning. *Hippocampus* **6**, 192–209 (1996).
- Wible, C. G. *et al.* Mnemonic correlates of unit activity in the hippocampus. *Brain Res.* **399**, 97–110 (1986).
- Eichenbaum, H., Kuperstein, M., Fagan, A. & Nagode, J. Cue-sampling and goal-approach correlates of hippocampal unit activity in rats performing an odour-discrimination task. *J. Neurosci.* **7**, 716–732 (1987).
- Wiener, S. I., Paul, C. A. & Eichenbaum, H. Spatial and behavioral correlates of hippocampal neuronal activity. *J. Neurosci.* **9**, 2737–2763 (1989).
- Deadwyler, S. A., Bunn, T. & Hampson, R. E. Hippocampal ensemble activity during spatial delayed-nonmatch-to-sample performance in rats. *J. Neurosci.* **16**, 354–372 (1996).
- Rolls, E. T., Cahusac, P. M. B., Feigenbaum, J. D. & Miyashita, Y. Responses of single neurons in the hippocampus of the macaque related to recognition memory. *Exp. Brain Res.* **93**, 299–306 (1993).
- Wilson, F. A. W., Riches, I. P. & Brown, M. W. Hippocampus and medial temporal cortex: Neuronal activity related to behavioral responses during the performance of memory tasks by primates. *Behav. Brain Res.* **40**, 7–28 (1990).
- Otto, T. & Eichenbaum, H. Neuronal activity in the hippocampus during delayed non-match to sample performance in rats: evidence for hippocampal processing in recognition memory. *Hippocampus* **2**, 323–334 (1992).
- Sakurai, Y. Involvement of auditory cortical and hippocampal neurons in auditory working memory and reference memory in the rat. *J. Neurosci.* **14**, 2606–2623 (1994).
- Young, B. J., Fox, G. D. & Eichenbaum, H. Correlates of hippocampal complex-spike activity in rats performing a nonspatial radial maze task. *J. Neurosci.* **14**, 6553–6563 (1994).
- Gothard, K. M., Skaggs, W. E., Moore, K. M. & McNaughton, B. L. Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task. *J. Neurosci.* **16**, 823–835 (1996).
- Gothard, K. M., Skaggs, W. E. & McNaughton, B. L. Dynamics of mismatch correction in the hippocampal ensemble code for space: interaction between path integration and environmental cues. *J. Neurosci.* **16**, 8027–8040 (1996).
- O'Keefe, J. & Burgess, N. Geometric determinants of the place fields of hippocampal neurons. *Nature* **381**, 425–428 (1996).
- Bunsey, M. & Eichenbaum, H. Conservation of hippocampal memory function in rats and humans. *Nature* **379**, 255–257 (1996).
- Morris, R. G. M. & Frey, U. Hippocampal synaptic plasticity: role in spatial learning or the automatic recording of attended experience? *Phil. Trans. R. Soc. Lond. B* **352**, 1489–1503 (1997).
- Tanila, H., Shapiro, M., Gallagher, M. & Eichenbaum, H. Brain aging: changes in the nature of information coding by the hippocampus. *J. Neurosci.* **17**, 5155–5166 (1997).

Acknowledgements. We thank J. O'Keefe for helping to design the experiment, and M. Burchinal for statistical advice. This work was supported by NIMH and NIA.

Correspondence and requests for materials should be addressed to H.E. (e-mail: hbe@bu.edu).

Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein

Pao-Tien Chuang & Andrew P. McMahon

Department of Molecular and Cellular Biology, The Biolabs, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA

The Hedgehog signalling pathway is essential for the development of diverse tissues during embryogenesis¹. Signalling is activated by binding of Hedgehog protein to the multipass membrane protein Patched (Ptc)^{2,3}. We have now identified a novel component in the vertebrate signalling pathway, which we name *Hip* (for *Hedgehog-interacting protein*) because of its ability to bind Hedgehog proteins. *Hip* encodes a membrane glycoprotein that binds to all three mammalian Hedgehog proteins with an affinity comparable to that of Ptc-1. *Hip*-expressing cells are located next to cells that express each *Hedgehog* gene. *Hip* expression is induced by ectopic Hedgehog signalling and is lost in *Hedgehog* mutants. Thus, *Hip*, like *Ptc-1*, is a general transcriptional target of Hedgehog signalling. Overexpression of *Hip* in cartilage, where *Indian hedgehog* (*Ihh*) controls growth⁴, leads to a shortened skeleton that resembles that seen when *Ihh* function is lost (B. St-Jacques, M. Hammerschmidt & A.P.M., in preparation). Our findings support a model in which *Hip* attenuates Hedgehog signalling as a result of binding to Hedgehog proteins: a negative regulatory feedback loop established in this way could thus modulate the responses to any Hedgehog signal.

All multicellular organisms require cell communication to regulate growth and differentiation in the embryo. One strategy for this is to establish discrete organizing centres that emit signals to coordinately control cell proliferation and cell fate determination. *Sonic hedgehog* (*Shh*), a member of the *Hedgehog* (*Hh*) family expressed in several vertebrate signalling centres, is a key signal in these patterning mechanisms^{1-3,5,6}. The other two mammalian Hh members, *Indian hedgehog* (*Ihh*) and *Desert hedgehog* (*Dhh*), are principally expressed in cartilage and testes, respectively^{4,7,8}, where they are essential for the development of these structures (refs 4, 8, and B. St-Jacques, M. Hammerschmidt & A.P.M., in preparation).

Genetic and biochemical analysis of the Hh signalling pathway has established that two membrane proteins encoded by the segment polarity genes *patched* (*ptc*) and *smoothened* (*smo*)^{2,3} are involved in ligand reception and signal transduction. Binding of Hh to Ptc relieves Ptc-mediated repression of Smo, thereby activating the signalling pathway. The targets transcriptionally activated in response to Hh signalling include *Ptc* itself, which creates a negative regulatory loop that may restrict further diffusion of Hh protein^{9,10}.

If we are to fully understand how Hh signalling is transformed into cellular responses, we must identify all the key components involved in the signalling process. As most components have been identified by genetic screening in the fruitfly *Drosophila*, we initiated a biochemical screen to identify novel regulatory factors required for transducing or modulating Hh signalling in vertebrates. We generated a biologically active fusion of the amino-terminal fragment of *Sonic hedgehog* (*Shh*-N) with alkaline phosphatase^{11,12} (*Shh*-N::AP) to allow colorimetric detection of cells which bound *Shh*. *Shh*-N::AP was used to screen a complementary DNA (cDNA) expression library from a mouse limb bud obtained on day 10 post-coitum. After screening around 60,000 clones, we identified one positive pool that promoted cell-surface binding to *Shh*-N::AP and binding in subsequent experiments to alkaline phosphatase-tagged

Ihh-N and *Dhh*-N (data not shown). From this pool, we isolated one positive clone containing a cDNA of 2.7 kilobases. As the protein encoded by the isolated clone binds to all three mammalian Hedgehog proteins, we have called this gene *Hedgehog-interacting protein* (*Hip*).

The complete *Hip* cDNA (2,669 base pairs) encodes a protein of 700 amino acids with a predicted relative molecular mass of 78,400 (Fig. 1a, c). The first 15 amino-acid residues of *Hip* constitute the first hydrophobic stretch, reminiscent of a signal peptide¹³, while the last 22 amino-acid residues are predicted to form a hydrophobic membrane-spanning domain. These are four potential N-linked glycosylation sites and two domains resembling epidermal growth factor (EGF)¹⁴ repeats that lie close to the carboxy terminus of *Hip* (Fig. 1a, c). The protein shares sequence similarity with the *Xenopus* Gene 5 product (21% sequence identity)¹⁵ and with two bacterial proteins, 345 and PCZA361.11 (15% and 16% sequence identity, respectively)¹⁶ (Fig. 1b). Comparison of the mouse *Hip* sequence

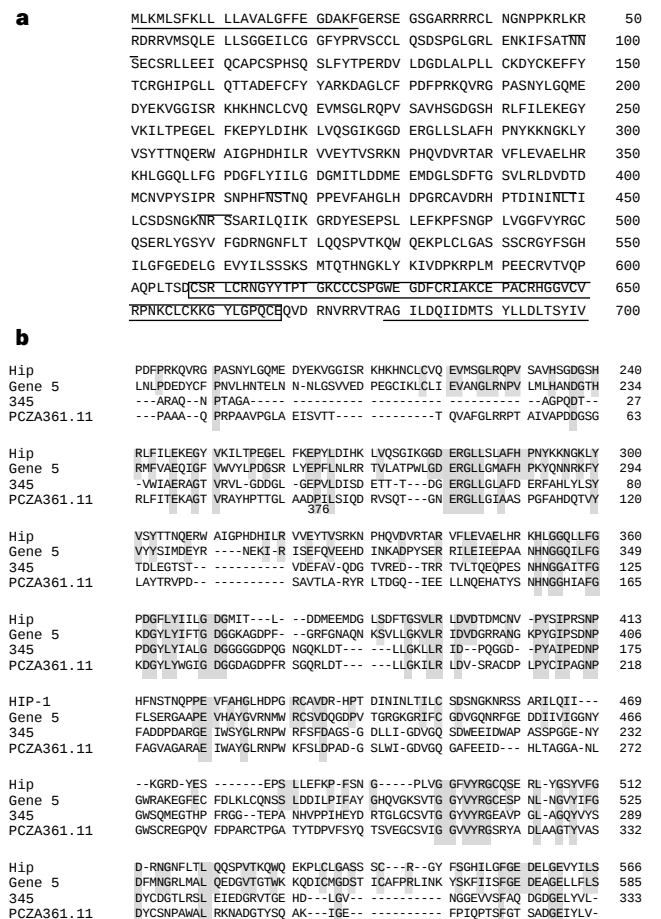


Figure 1 Sequence analysis of the *Hip* cDNA. **a**, Predicted translation product of the *Hip* cDNA. The hydrophobic stretches at the N- and C-termini are underlined, two EGF-like domains are boxed and four potential N-linked glycosylation sites are overlined. **b**, Amino-acid sequence alignment between *Hip*, Gene 5, 345 and PCZA361.11. Only regions with significant sequence identity are shown. Shaded regions represent sequences identical in all four proteins as well as sequences identical between only *Hip* and Gene 5. Numbers represent the amino-acid position in the corresponding protein. **c**, Schematic diagram showing the structure of *Hip* protein. I, hydrophobic stretch at the N terminus; II, EGF-like domains; III, hydrophobic stretch at the C terminus. Four potential N-linked glycosylation sites are represented by 'Y'.

with that of zebrafish, chick and human (data not shown) indicates that the *Xenopus Gene 5* is not the *Xenopus Hip* orthologue.

The finding that Hip can cause Hedgehog::AP fusion proteins to be retained on the cell surface and the presence of a putative signal peptide and transmembrane domain strongly suggest that it is a membrane-associated protein. Consistent with this, when expressed in cultured cells, a Myc-epitope tagged form of Hip (Myc-Hip) associated preferentially with the cell pellet fraction (Fig. 2a, lane 2) and was absent from the supernatant (Fig. 2a, lane 1). As the last 22 amino-acid residues encode a hydrophobic region, we used a Hip protein lacking these residues (Myc-Hip Δ C22) to test the possibility that this sequence might function as a transmembrane domain. A significant fraction of Myc-Hip Δ C22 was detected in the supernatant, indicating that this domain is essential for membrane binding (Fig. 2a, lane 3). To rule out the possibility that Hip is anchored to the membrane through a glycosylphosphatidylinositol (GPI)-linkage, we tested its susceptibility to cleavage by phosphatidylinositol phospholipase C (PI-PLC). Whereas a control GPI-linked protein, RETL1 (ref. 17), was cleaved by PI-PLC and released into the medium (Fig. 2a lane 8), Myc-Hip remained associated with the cell (Fig. 2a, compare lanes 12 and 13). Thus, Hip is a type I transmembrane protein¹⁸ with its last 22 amino-acid residues involved in membrane anchoring. The presence of four potential N-linked glycosylation sites suggests that Hip is a glycoprotein and, consistent with this view, treatment with endoglycosidase F to remove N-linked glycosylation produced a protein that migrated faster on SDS-polyacrylamide gel electrophoresis (Fig. 2a, lanes 14, 15). To determine whether Hip binds Shh-N directly as expected, we transfected cultured cells with the secreted form of Hip (Myc-Hip Δ C22), and then immunoprecipitated this from the medium with a Shh-N::IgG fusion protein (Fig. 2a, lane 5). Thus, Hip does interact directly with Shh-N.

Ptc-1 has been reported to confer cell-surface binding of Shh^{19,20}. Using the cell-surface-binding assay, we found that Ptc-1, like Hip, also binds to all three mammalian Hedgehog proteins (data not shown). The dissociation constants (K_d) for Shh-N::AP binding to Hip and Ptc-1 were similar, approximately 5 nM and 4 nM, respectively (Fig. 2b). The finding that the K_d for Ptc-1 binding to Shh-N::AP is higher than the published value for Shh-N^{19,20} (0.5–1.0 nM) probably reflects a weakened interaction of the biologically less active fusion protein¹².

We examined the temporal and spatial expression pattern of *Hip* in mouse embryos from 7.5 to 14.5 days post-coitum (dpc). When *Shh* and *Ptc-1* were first detected at 7.75 dpc^{21,22}, *Hip-1* expression was undetectable (data not shown). Subsequently, *Shh* expression was detected in several signalling centres, including the notochord, floor plate and zone of polarizing activity (ZPA) of the limb and in several endodermal derivatives^{7,21}. In the caudal neural tube at 8.75 dpc, *Shh* expression, which is restricted to the notochord, is implicated in ventralizing the somites and neural tube. Here, *Hip* was activated at the ventral midline of the neural tube and in the ventral medial somites (Fig. 3a, b), next to *Shh*-expressing cells but anterior to where *Ptc-1* is first upregulated (data not shown). The close association of *Hip* and *Shh* expression in the notochord was maintained in the caudal region at tail bud stages (Fig. 3c, d). When *Shh* was induced in the floor plate at the ventral midline of the more rostral neural tube, *Hip* expression was lost in the floor plate but retained in the ventral half of the neural tube and in the sclerotome of the adjacent somites (Fig. 3a, b, e, f). In the midbrain, where *Shh* is implicated in the induction of dopaminergic neurons ventrally¹, *Hip* expression was confined to two lateral stripes immediately adjacent to *Shh*-expressing cells in the floor plate (Fig. 3e–h). *Hip* was also expressed in gut mesenchyme along the length of gastrointestinal tract next to *Shh*-expressing cells (Fig. 3g, h) and in

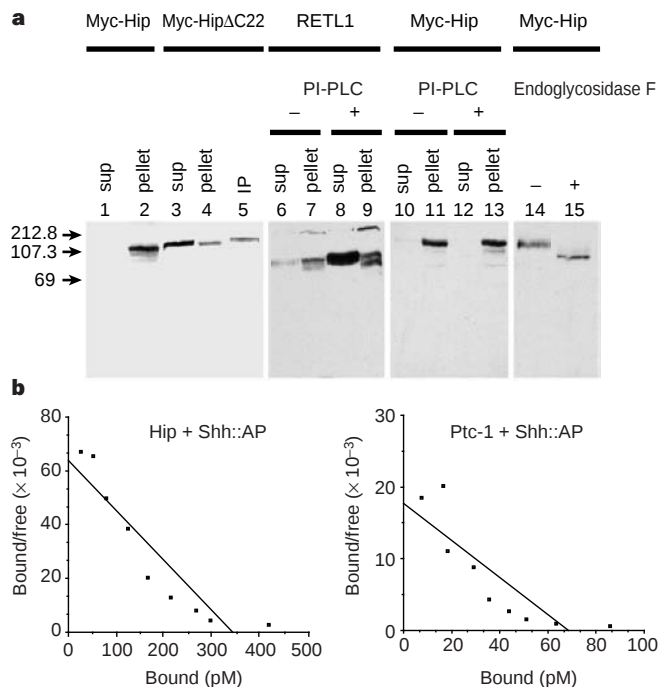


Figure 2 Analysis of the Hip protein. **a**, Lanes 1–5, western blot of the cell pellet and concentrated supernatant (sup) from COS-7 cells transfected with Myc-tagged Hip or Myc-Hip Δ C22, probed with an anti-Myc antibody. Whereas Myc-Hip is almost exclusively associated with the cell pellet fraction (lane 2), a significant portion of Myc-Hip Δ C22 is present in the medium (lane 3). Myc-Hip Δ C22 in the medium can be immunoprecipitated with Shh-N::IgG (lane 5). Lane 6–13, western blots of the cell pellet and supernatant from COS-7 cells transfected with RETL1 or Myc-Hip probed with an anti-RETL1 or anti-Myc antibody. As a result of PI-PLC treatment, a significant portion of RETL1, a GPI-

linked protein, is released into the medium (lane 8). In contrast, Myc-Hip remains associated with the cell pellet fraction (lane 13). Lanes 14, 15, western blot of Hip protein in the absence or presence of endoglycosidase F. Treatment with endoglycosidase F results in a significantly faster migration (lane 15) on SDS-PAGE compared with untreated cells (lane 14). The location of protein size standards ($\times 10^{-3}$) are indicated by the arrows on the left. **b**, Scatchard analysis¹¹ of the dissociation constants (K_d) between Shh and two Hedgehog-binding proteins, Hip and Ptc-1.

mesenchyme of the posterior half of the limb bud anterior to *Shh* expression in the ZPA (Fig. 3e, f). These results suggest that *Hip* is well placed to participate in *Shh* signalling. Further, as activation of *Hip* follows upregulation of *Ptc-1*, *Hip* may be a transcriptional target of *Shh* signalling, activated shortly after the initial response to *Shh*. Later in development, *Shh* is expressed in the epithelia of a wide variety of tissues⁷, for instance the epithelium of the lung (Fig. 3i), gut (Fig. 3m), whisker (Fig. 3o), hair and rugae. In all cases, *Hip* was expressed in the underlying mesenchyme (Fig. 3j, n, p, and data not shown).

Expression of *Hip* was also detected in close association with sites of *Ihh* and *Dhh* expression. *Ihh* is expressed in the prehypertrophic chondrocytes, where it acts as a chondrocyte mitogen and controls the rate of chondrocyte differentiation (refs 4, 7, and B. St-Jacques, M. Hammerschmidt & A.P.M., in preparation). We found that *Hip*, like *Ptc-1*, was expressed in the perichondrium, including regions flanking *Ihh* expression in the appendicular (Fig. 3k) and axial skeleton (Fig. 3l). *Dhh* is expressed in Sertoli cells of the testis, where it is required for male germline development⁸. *Hip* and *Ptc-1* were expressed in the androgen-producing interstitial somatic cells (the

Leydig cells) which are thought to respond to *Dhh* signalling (Fig. 4o and ref. 8). Thus, *Hip*, like *Ptc-1*, may be a transcriptional target of all mammalian Hh pathways and may have a general role in modulating all Hedgehog signalling.

We examined the regulation of *Hip* expression in transgenic mouse embryos expressing *Shh* ectopically, or in embryos carrying mutant *Hedgehog* genes. When *Shh* was ectopically expressed in the dorsal spinal cord²¹ (arrows in Fig. 4b, f, compare with Fig. 4a, e), *Hip* (arrows in Fig. 4d, h, compare with Fig. 4c, g) and *Ptc-1* (arrow in Fig. 4j, compare with Fig. 4i) were ectopically expressed in regions immediately dorso-lateral to ectopic *Shh* expression. Further, expression of a transgene encoding a dominant-negative form of cyclic AMP-dependent protein kinase A in the dorsal neural tube, which ectopically activates Hh targets such as *Ptc-1*²³, also led to ectopic *Hip* expression (arrow in Fig. 4l). Finally, in *Shh* (Fig. 4n) and *Dhh*⁸ mutants (Fig. 4p), *Hip* expression was absent in Hh-responsive cells (Fig. 4m, o). We conclude that *Hip*, like *Ptc-1*, is a general transcriptional target of Hh signalling.

The fact that *Hip* binds Hedgehog proteins directly and that its expression requires Hh signalling suggests that *Hip* may modify Hh

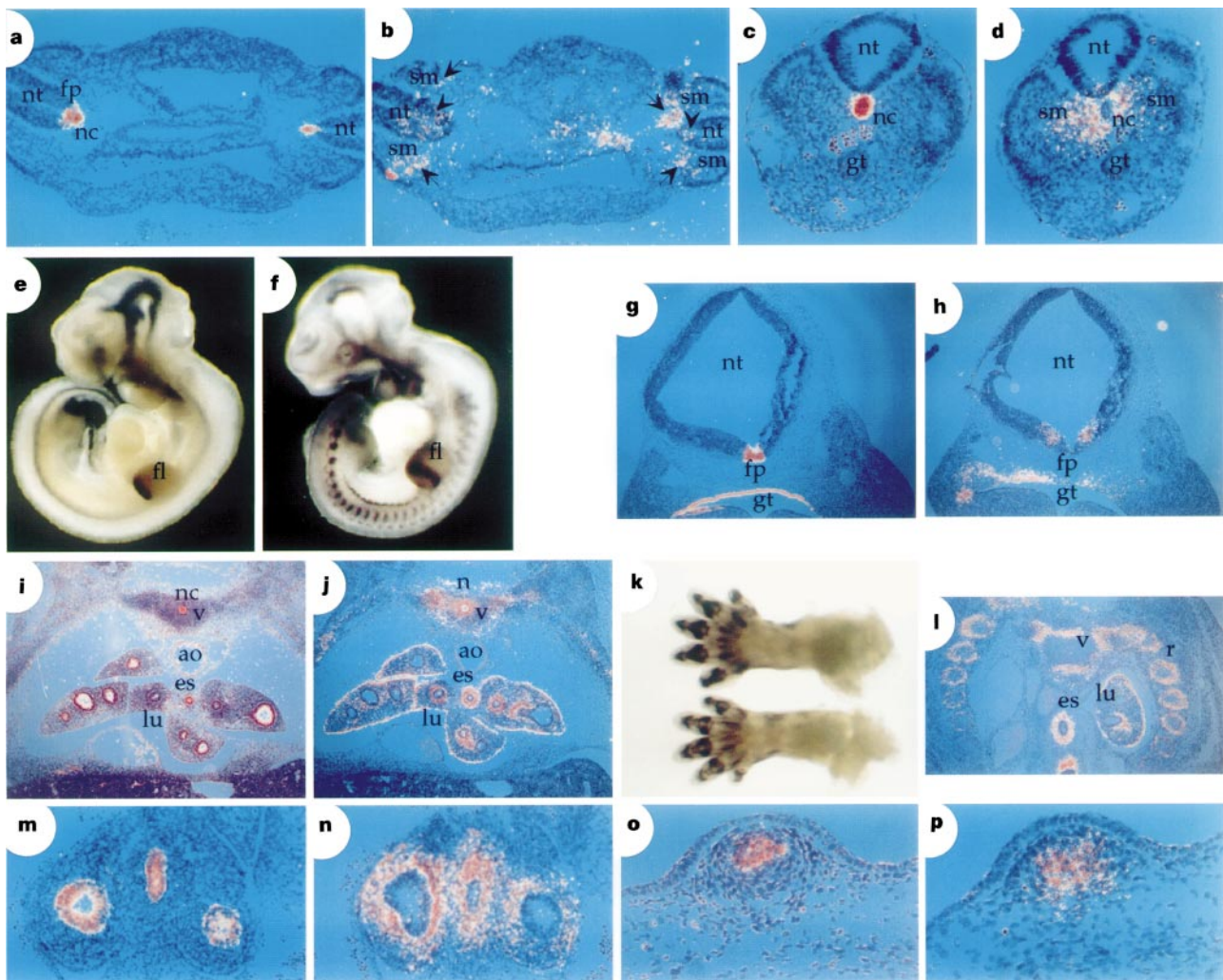


Figure 3 Expression of *Hip* during embryogenesis. See text for details of expression pattern. **e, f, k**, Whole-mount *in situ* hybridization. **a-d, g-j, l-p**, *In situ* hybridization performed on sectioned tissue. **a, c, e, g, i, m, o**, *In situ* hybridization to mouse *Shh* messenger RNA (mRNA). **b, d, f, h, j, k, l, n, p**, *In situ* hybridization to mouse *Hip* mRNA. **a/b, c/d, g/h, i/j, m/n, o/p**, Each pair represents adjacent sections. **a, b**, Section through a 9.0-dpc embryo. Arrowheads indicate *Hip* expression in the neural tube and arrows point to *Hip* expression in the somite. **c, d**, Section through a 10.5-dpc embryo. **e, f**, Lateral view of 10.5-dpc embryos. **g, h**,

Transverse section through the midbrain of a 10.5-dpc embryo. **i, j**, Section through the thoracic region of a 14.5-dpc embryo. **k**, Dorsal view of forelimb dissected at 14.5 dpc. **l**, Transverse section through the rib cage of a 14.5-dpc embryo. **m, n**, Transverse section through the gut of a 14.5-dpc embryo. **o, p**, Section through the skin of a 14.5-dpc embryo. ao, Aorta; es, esophagus; fl, forelimb; fp, floor plate; gt, gut; lu, lung; nc, notochord; sm, somite; nt, neural tube; r, rib; v, vertebral body.

signalling. *Hip* could function either in a negative manner to attenuate a Hh signal or positively to facilitate signal transduction. In the first model, overexpression of *Hip* is likely to result in loss of Hh signalling by competition with Ptc for Hh binding. In the second model, *Hip* overexpression may enhance signal transduction, perhaps by facilitating Hh protein diffusion or increasing the affinity of Hh for other proteins on the membrane such as Ptc. To distinguish between these models, we expressed *Hip* ectopically in the devel-

oping endochondral skeleton using a rat $\alpha 1(\text{II})$ collagen promoter/enhancer²⁴. $\alpha 1(\text{II})$ collagen is expressed in proliferating chondrocytes before hypertrophic differentiation²⁵. Loss-of-function studies in the mouse have demonstrated that *Ihh* is essential for proliferation of chondrocytes (B. St-Jacques, M. Hammerschmidt & A.P.M., in preparation). *Ihh* mutants displayed a markedly shortened skeleton, most pronounced in the limbs (compare Fig. 5a and 5c). Interestingly, transgenic animals expressing the *Hip* transgene (*Hip*Tg) showed severe skeletal defects, including short-limbed dwarfism (Fig. 5b) remarkably similar to that observed in *Ihh* mutants. Histological analysis of *Hip*Tg animals revealed a marked reduction in the zone of undifferentiated chondrocytes (compare Fig. 5d, e and arrows in Fig. 5g, h), similar to *Ihh* mutants (Fig. 5f and arrows in Fig. 5i) but opposite to the expansion of proliferating, undifferentiated chondrocytes observed in animals overexpressing *Ihh*⁴. In addition, ectopic calcification occurred in similar regions of the skeleton as those observed in *Ihh* mutants (B. St-Jacques, M. Hammerschmidt & A.P.M., in preparation) and *Hip*Tg animals (data not shown). Finally, *Ptc-1* expression in proliferating chondrocytes was greatly reduced in *Hip*Tg animals whereas *Ihh* expression in prehypertrophic chondrocytes was unaltered (data not shown). Thus, overexpression of *Hip* diminishes *Ihh* signalling, consistent with a general model in

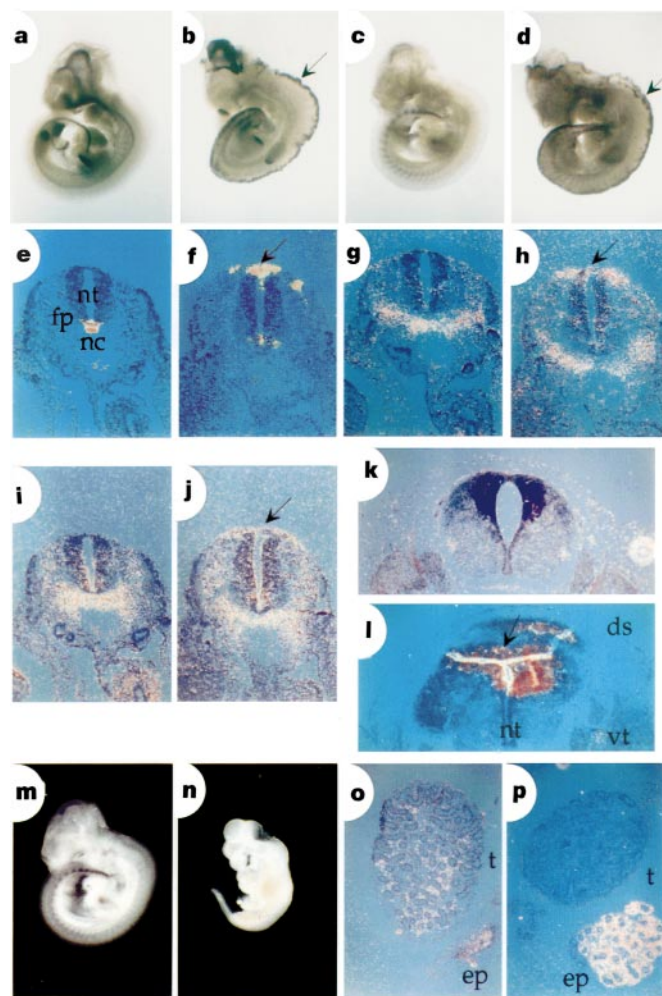


Figure 4 Expression of *Hip* transcripts in animals with ectopic or defective Hedgehog signalling. **a-d**, **m**, **n**, Whole-mount; **e-l**, **o**, **p**, section *in situ* hybridization. **a**, **b**, **e**, *In situ* hybridization to mouse *Shh* mRNA. **f**, *In situ* hybridization to chick *Shh* mRNA. **c**, **d**, **g**, **h**, **k**, **l**, **m**, **n**, **o**, **p**, *In situ* hybridization to mouse *Hip* mRNA. **i**, **j**, *In situ* hybridization to mouse *Ptc-1* mRNA. **a-d**, Lateral view of two 10.5-dpc wild-type (**a**, **c**) and two transgenic animals (**b**, **d**) expressing mouse *Shh* under the control of a *Wnt-1* enhancer. **e-j**, Section through the spinal cord of wild-type (**e**, **g**, **i**) and transgenic (**f**, **h**, **j**) embryos (10.5 dpc) expressing chick *Shh* under *Wnt-1* control. Ectopic *Hip* (arrows in **d** and **h**) and *Ptc-1* (arrow in **j**) expression is seen in the dorsal neural tube where *Shh* is ectopically expressed (arrows in **b** and **f**). **k**, **l**, Section through the spinal cord at 12.5 dpc of a wild-type (**k**) and transgenic embryo (**l**) carrying a transgene in which dominant-negative PKA is under *Wnt-1* control. Ectopic *Hip* expression (arrow in **l**) is seen in the dorsal neural tube where the Hh pathway is ectopically activated. **m**, **n**, Lateral view of a 10.5-dpc wild-type (**m**) and *Shh* mutant (**n**) embryo in which *Hip* expression is lost. **o**, **p**, Section through the testes of a wild-type (**o**) and *Dhh* mutant animal (**p**) (3 weeks *post partum*). Whereas *Hip* is expressed in the Leydig cells of wild-type testes (**o**), its expression is lost in a *Dhh* mutant (**p**). Mesenchymal expression of *Hip* is maintained in the epididymis, where *Shh* expression in the epithelium of *Dhh* mutant testes (**p**) is unaltered. nt, Neural tube; fp, floor plate; nc, notochord; ds, dorsal; vt, ventral; ep, epididymis; t, testis.

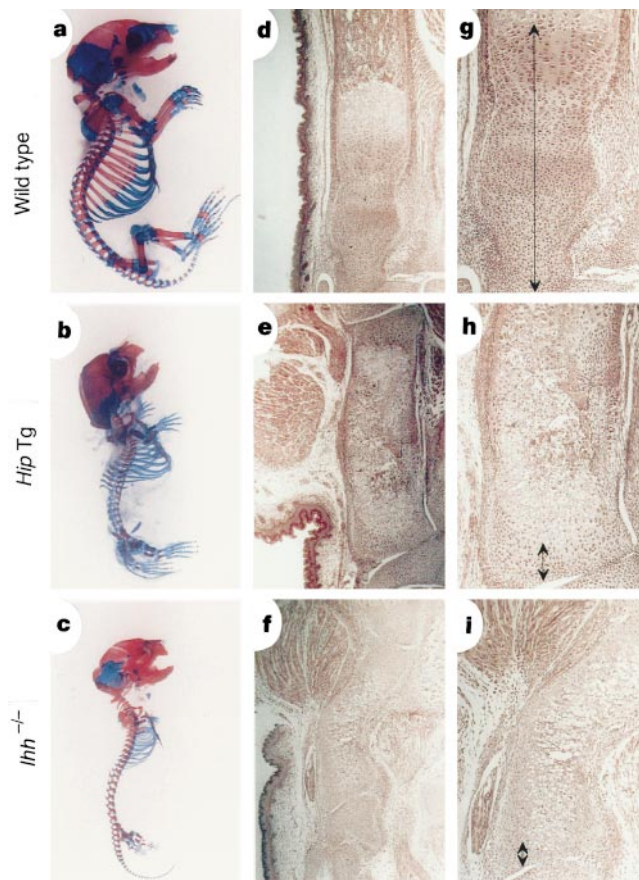


Figure 5 Analysis of skeleton in animals overexpressing *Hip* in chondrocytes. **a-c**, Skeleton of a wild-type (**a**), *Hip* transgenic (*Hip*Tg) (**b**) and *Ihh* mutant (**c**) 18.5-dpc embryo stained with Alcian blue and Alizarin red. Short-limbed dwarfism with general reduction of Alizarin red S staining is seen in the transgenic embryo, similar to that observed in *Ihh* mutants. **d-i**, Haematoxylin- and eosin-stained sections through the radius of a wild-type (**d**, **g**), *Hip*Tg (**e**, **h**) and *Ihh* mutant (**f**, **i**) 18.5-dpc embryo. A marked reduction of immature chondrocytes is observed in *Hip*Tg animals (arrows in **h**) and *Ihh* mutants (arrows in **i**) compared with that seen in the wild type (arrows in **g**). Photographs of **d-f** and **g-i** were taken at the same magnification, respectively.

which upregulation of *Hip* by Hedgehog signals attenuates further Hedgehog signalling. Attenuation of signalling by physical binding of ligand may then allow differential responses to be generated within a field of competent cells. In *Drosophila*, a single *Hedgehog* gene mediates all Hedgehog signalling whereas in mammals three *Hedgehog* genes have largely non-overlapping activities. We have failed to identify a *Hip* orthologue in *Drosophila*. This suggests that in addition to diversification of Hedgehog signals, vertebrates may have evolved novel mechanisms of Hedgehog signal modulation that could facilitate distinct aspects of vertebrate development. □

Methods

Standard molecular biology and protein biochemistry techniques were performed as described²⁶.

Database searches. The *Hip* protein sequence was used to search the National Center for Biotechnology Information database using the BLAST program. Motif analysis of *Hip* and sequence alignments from *Hip*, Gene 5, 345 and pCZA361.11 were done with the GeneWorks program (IntelliGenetics). Transmembrane prediction was performed by SOSUI algorithm²⁷.

Epitope tagging, co-immunoprecipitation, endoglycosidase and PI-PLC treatment. To generate Myc–*Hip*, *Hip* cDNA was cloned into pCDNA3 (Invitrogen) and a Myc epitope was inserted between amino-acid residues 23 and 24. To generate Myc–*Hip*ΔC22, a stop codon was introduced 5' to the last 22 residues in Myc–*Hip*. Shh–N::IgG was generated by cloning a DNA fragment encoding Shh–N into the pCD5IgG1 vector. To determine the localization of Myc–*Hip* and Myc–*Hip*ΔC22, COS-7 cells transfected with Myc–*Hip* or Myc–*Hip*ΔC22 using LipofectAMINE were collected and lysed 48 h after transfection. The supernatant was filtered and concentrated by precipitation with trichloroacetic acid (TCA). To immunoprecipitate Myc–*Hip*ΔC22, unconcentrated supernatant from COS-7 cells transfected with Myc–*Hip*ΔC22 was incubated with Shh–N::IgG bound to Protein A beads for 4 h at 4 °C. After extensive washing with buffer (25 mM HEPES (pH 7.6), 0.1 mM EDTA, 100 mM NaCl and 0.1% NP40), the beads were resuspended in sample buffer. Protein was run on SDS–PAGE and western blotting was done using anti-Myc antibody. To examine glycosylation of *Hip*, COS-7 cells transfected with Myc–*Hip* were lysed in RIPA buffer 48 h after transfection. The supernatant was mixed with an equal volume of buffer (20 mM sodium phosphate (pH 6.8), 40 mM EDTA, 0.1% SDS and 20 mM β-mercaptoethanol), boiled for 10 min and 0.4 U of endoglycosidase F (Boehringer Mannheim) was then added. The mixture was incubated at 37 °C for 24 h and analysed by western blotting using an anti-Myc antibody.

In situ hybridization. Whole-mount *in situ* hybridization using digoxigenin-labelled probes and section *in situ* hybridization using ³⁵S-labelled probes were performed as described²⁸.

Generation of transgenic animals. To mis-express *Hip* under the α1(II) collagen promoter/enhancer, the full-length *Hip* cDNA was cloned into the α1(II) collagen expression vector. Transgenic animals were generated by pronuclear injection as described²⁹. Staining of skeleton for bone and cartilage was as described²⁹.

Received 23 November; accepted 18 December 1998.

- Hammerschmidt, M., Brook, A. & McMahon, A. P. The world according to hedgehog. *Trends Genet.* **13**, 14–21 (1997).
- Tabin, C. J. & McMahon, A. P. Recent advances in Hedgehog signaling. *Trends Cell Biol.* **7**, 442–446 (1997).
- Ingham, P. W. Transducing Hedgehog: the story so far. *EMBO J.* **17**, 3505–3511 (1998).
- Vortkamp, A. et al. Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. *Science* **273**, 613–622 (1996).
- Jessell, T. M. & Dodd, J. Floor plate-derived signals and the control of neural cell pattern in vertebrates. *Harvey Lect.* **86**, 87–128 (1990).
- Schwabe, J. W. R., Rodriguez-Esteban, C. & Izpisua Belmonte, J. C. Limbs are moving: where are they going? *Trends Genet.* **14**, 229–235 (1998).
- Bitgood, M. J. & McMahon, A. P. Hedgehog and Bmp genes are coexpressed at many diverse sites of cell-cell interaction in the mouse embryo. *Dev. Biol.* **172**, 126–138 (1995).
- Bitgood, M. J., Shen, L. & McMahon, A. P. Sertoli cell signaling by Desert hedgehog regulates the male germline. *Curr. Biol.* **6**, 298–304 (1996).
- Porter, J. A. et al. Hedgehog patterning activity: role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. *Cell* **86**, 21–34 (1996).
- Chen, Y. & Struhl, G. Dual roles for patched in sequestering and transducing Hedgehog. *Cell* **87**, 553–563 (1996).
- Cheng, H. J. & Flanagan, J. G. Identification and cloning of ELF-1, a developmentally expressed ligand for the Mek4 and Ser receptor tyrosine kinases. *Cell* **79**, 157–168 (1994).
- Yang, Y. et al. Relationship between dose, distance and time in Sonic Hedgehog-mediated regulation of anteroposterior polarity in the chick limb. *Development* **124**, 4393–4404 (1997).

- von Heijne, G. Protein targeting signals. *Curr. Opin. Cell Biol.* **2**, 604–608 (1990).
- Muskavitch, M. A. & Hoffmann, F. M. Homologs of vertebrate growth factors in *Drosophila melanogaster* and other invertebrates. *Curr. Top. Dev. Biol.* **24**, 289–328 (1990).
- Brown, D. D. et al. The thyroid hormone-induced tail resorption program during *Xenopus laevis* metamorphosis. *Proc. Natl Acad. Sci. USA* **93**, 1924–1929 (1996).
- Schneider, D., Bruton, C. J. & Chater, K. F. Characterization of spaA, a *Streptomyces coelicolor* gene homologous to a gene involved in sensing starvation in *Escherichia coli*. *Gene* **177**, 243–251 (1996).
- Sanicola, M. et al. Glial cell line-derived neurotrophic factor-dependent RET activation can be mediated by two different cell-surface accessory proteins. *Proc. Natl Acad. Sci. USA* **94**, 6238–6243 (1997).
- Matlack, K. E., Mothes, W. & Rapoport, T. A. Protein translocation: tunnel vision. *Cell* **92**, 381–390 (1998).
- Stone, D. M. et al. The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* **384**, 129–134 (1996).
- Marigo, V., Davey, R. A., Zuo, Y., Cunningham, J. M. & Tabin, C. J. Biochemical evidence that Patched is the Hedgehog receptor. *Nature* **384**, 176–179 (1996).
- Echelard, Y. et al. Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* **75**, 1417–1430 (1993).
- Goodrich, L. V., Johnson, R. L., Milenkovic, L., McMahon, J. A. & Scott, M. P. Conservation of the hedgehog/patched signaling pathway from flies to mice: induction of a mouse patched gene by Hedgehog. *Genes Dev.* **10**, 301–312 (1996).
- Epstein, D. J., Marti, E., Scott, M. P. & McMahon, A. P. Antagonizing cAMP-dependent protein kinase A in the dorsal CNS activates a conserved Sonic hedgehog signaling pathway. *Development* **122**, 2885–2894 (1996).
- Nakata, K. et al. Osteoarthritis associated with mild chondrodysplasia in transgenic mice expressing alpha 1(I) collagen chains with a central deletion. *Proc. Natl Acad. Sci. USA* **90**, 2870–2874 (1993).
- Cheah, K. S., Lau, E. T., Au, P. K. & Tam, P. P. Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilage during development. *Development* **111**, 945–953 (1991).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
- Hirokawa, T., Boon-Chieng, S. & Mitaku, S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**, 378–379 (1998).
- Wilkinson, D. G. & Nieto, M. A. Detection of messenger RNA by *in situ* hybridization to tissue sections and whole mounts. *Methods Enzymol.* **225**, 361–373 (1993).
- Hogan, B., Beddington, R., Costantini, F. & Lacy, E. *Manipulating the Mouse Embryo: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1994).

Acknowledgements. We thank H.-J. Cheng for advice on expression cloning, J.-L. Chen for help with sequencing, B. St-Jacques for *Ihh* mutant embryos, S. Lee, J. McMahon and M. Bitgood for performing *in situ* hybridization, N. Wu for pronuclear injection, D. Faria and B. Klumpar for histological section, M. Scott for the *Patched-1* expression clone, B. Seed for pCD5IgG1 vector, Y. Yamada for the α1(II) collagen expression vector and M. Sanicola for the RETL1 clone, anti-RETL1 antibody and a protocol for PI–PLC cleavage. We thank past and present members of the McMahon lab for discussion and critical reading of the manuscript. P.-T.C. is a fellow of the Leukemia Society of America. This work was supported by a grant from the NINDS to A.P.M. The GenBank accession number for *Hip* is AF116865.

The Rab5 effector EEA1 is a core component of endosome docking

Savvas Christoforidis*, Heidi M. McBride*, Robert D. Burgoyne† & Marino Zerial*

* European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69117 Heidelberg, and Max Planck Institute for Molecular Cell Biology and Genetics, Pfotenhauerstrasse, Dresden, 01307 Germany

† The Physiological Laboratory, University of Liverpool, Crown Street, Liverpool L69 3BX, UK

Intracellular membrane docking and fusion requires the interplay between soluble factors and SNAREs. The SNARE hypothesis¹ postulates that pairing between a vesicular v-SNARE and a target membrane z-SNARE is the primary molecular interaction underlying the specificity of vesicle targeting as well as lipid bilayer fusion. This proposal is supported by recent studies using a minimal artificial system². However, several observations demonstrate that SNAREs function at multiple transport steps and can pair promiscuously, questioning the role of SNAREs in conveying vesicle targeting^{3–6}. Moreover, other proteins have been shown to be important in membrane docking or tethering^{7–9}. Therefore, if the minimal machinery is defined as the set of proteins sufficient to reproduce *in vitro* the fidelity of vesicle targeting, docking and fusion as *in vivo*, then SNAREs are not sufficient to specify vesicle targeting. Endosome fusion also requires cytosolic factors and is regulated by the small GTPase Rab5 (refs 10–20). Here we show that Rab5-interacting soluble proteins can completely substitute for cytosol in an *in vivo*

endosome-fusion assay, and that the Rab5 effector EEA1 is the only factor necessary to confer minimal fusion activity. Rab5 and other associated proteins seem to act upstream of EEA1, implying that Rab5 effectors comprise both regulatory molecules and mechanical components of the membrane transport machinery. We further show that EEA1 mediates endosome docking and, together with SNAREs, leads to membrane fusion.

Rab proteins and their downstream effectors contribute to vesicle targeting specificity¹⁰ and their delivery to the membrane has been proposed to regulate the assembly of SNARE complexes^{11–14}. Early-endosome fusion requires cytosol, Rab5 (ref. 15), its effector Rabaptin-5 (refs 16, 17) and several cytosolic and peripheral membrane proteins^{18–20}, which modulate the membrane fusion activity. However, it is not clear to what extent these molecules alone can support endosome fusion. For example, the Rabaptin-5/Rabex-5 complex is necessary but not sufficient to replace cytosol in the fusion reaction¹⁶.

To determine the minimal factors required for this reaction, we sought to purify all possible Rab5 effectors and determine their contribution to endosome docking and fusion. For this we used an affinity-chromatography approach. We found that 22 proteins from bovine brain cytosol bound specifically to immobilized glutathione *S*-transferase (GST)–Rab5 (Fig. 1a). This specificity is apparent from five lines of evidence. (1) Most proteins selectively bound the GTP- γ S form of Rab5. Only two proteins were enriched in the Rab5:GDP eluate and, owing to incomplete nucleotide exchange, these were also present in the Rab5:GTP- γ S column. (2) Immobilized GST alone (Fig. 1a) or GST–Rab4 (not shown) yielded a distinct protein pattern. (3) As expected, we purified the previously described Rab5 effectors, Rabaptin-5 and Rabaptin-5 β , their associated co-factor Rabex-5, and identified EEA1, as described recently¹⁸. (4) Abundant cytosolic proteins such as actin or XMAP215 were not detected in the eluate (not shown). (5) These proteins displayed activity in endosome fusion, as we show later.

We next tested whether the eluate of the Rab5 affinity column could functionally substitute for cytosol in the *in vitro* assay of

homotypic endosome fusion. Basal fusion reactions were performed with cytosol at a final concentration of 3 mg ml^{–1}, which corresponds to 80% of the maximal fusion efficiency obtained at saturating cytosol concentration (5 mg ml^{–1}). Whereas the GST–Rab5:GDP-column eluate or a control protein (GST, not shown) yielded only background fusion, the eluate of the GST–Rab5:GTP- γ S column efficiently supported endosome fusion (Fig. 1b). Remarkably, 0.4 μ g of the eluate, in which the known Rab5 effectors (EEA1, Rabaptin-5 and Rabex-5) were present at a concentration similar to cytosol (data not shown), was able to support fusion to cytosolic levels. Moreover, the eluate stimulated fusion in a concentration-dependent manner, reaching 50% of the maximal fusion signal as measured by solubilization of the membranes with detergent. It should be noted that *N*-ethylmaleimide-sensitive fusion protein (NSF), which is necessary for endosome fusion^{21,22} was present on the membrane (not shown). These results indicate that the Rab5-interacting proteins can account for the cytosolic requirement in endosome docking and fusion.

Given the multitude of polypeptides eluted from the Rab5:GTP- γ S column, we investigated their relative contribution in the *in vitro* assay. We subjected the eluted proteins to further separation by Superose-6 gel-filtration chromatography (Fig. 2a). Judging from the protein elution profile, the sum of fractions 19, 27, 32, 36 and 39 comprises the entire polypeptide composition of the column eluate. These fractions were tested in early-endosome fusion either

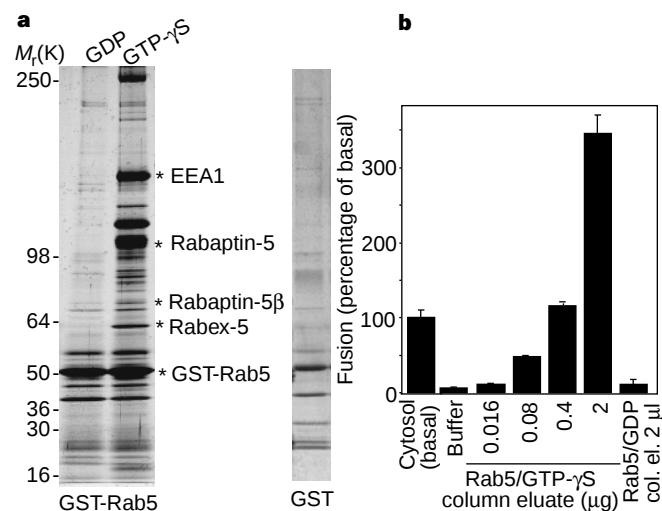


Figure 1 Twenty-two cytosolic proteins bind to Rab5 and can substitute cytosol in endosome fusion. **a**, Left panel: GST–Rab5:GDP and GST–Rab5:GTP- γ S affinity columns were incubated with bovine brain cytosol and bound proteins were eluted and analysed by gradient SDS–polyacrylamide gel electrophoresis (PAGE) (6–17%) followed by silver staining. Molecular mass standards are indicated on the left side of the gel. EEA1, Rabaptin-5, Rabaptin-5 β , Rabex-5 and Rab5:GST are all indicated, as identified by western blot and microsequencing (not shown). Right panel: GST column was treated as described above for GST–Rab5. **b**, The endosome homotypic fusion assay was done with 3 mg ml^{–1} cytosol (basal), buffer, increasing concentration of Rab5:GTP- γ S column eluate or Rab5:GDP column eluate.

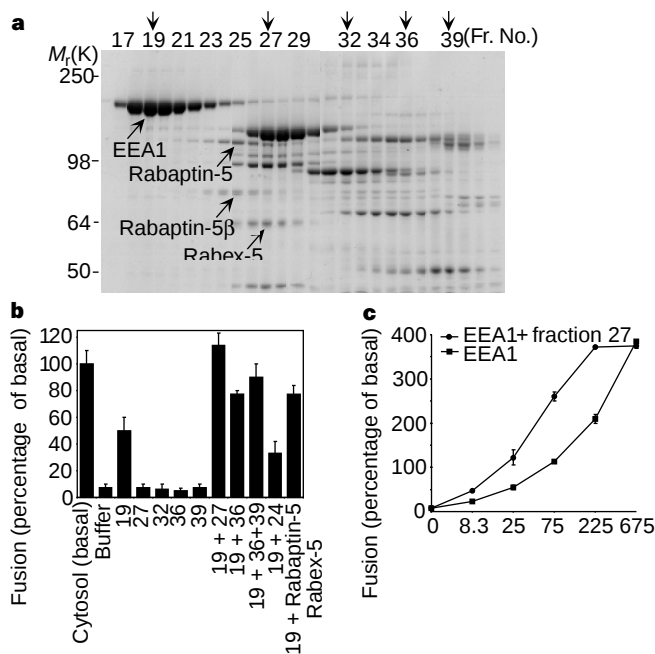


Figure 2 Fractionated eluate reveals a core activity of EEA1, regulated by the other effectors. **a**, The mixture of the Rab5 effectors was separated on a Superose 6 gel filtration column. Fractions eluted from the column were analysed by SDS–PAGE followed by Coomassie staining. Fractions 19, 27, 32, 36 and 39 (top arrows) were tested in the endosome fusion assay (see below). Molecular mass standards are indicated on the left. Arrows mark the position of EEA1, Rabaptin-5, Rabaptin-5 β and Rabex-5 identified by western blotting (not shown). **b**, Early-endosome fusion assay done in the presence of cytosol (basal) or the reagents indicated. Fraction 19 was added to 25 nM final concentration of EEA1 in the fusion reaction. Fraction 27 was added in amounts (3 μ l) that give a final concentration of Rabaptin-5 equal to that obtained by the addition of cytosol. The other fractions were added at equal volumes to fraction 27. Rabaptin-5/Rabex-5 complex was added at a concentration similar to that of fraction 27 (20 nM). **c**, *In vitro* fusion between early endosomes was done in the presence or absence of fraction 27 with increasing amounts of EEA1. Fraction 27 was added at a final concentration of Rabaptin-5 equal to that found in cytosol (3 μ l). Fusion obtained with 3 mg ml^{–1} cytosol is normalized to 100 %.

individually or in combination. Fractions 27, 32, 36 and 39 could not support endosome fusion (Fig. 2b). Unexpectedly, fraction 19, which contained almost exclusively a single protein of relative molecular mass 170,000 (M_r 170K), yielded significant fusion activity. We identified this protein as EEA1 by western blotting and mass spectrometry analysis (not shown).

When the endosomes were supplied with EEA1 at endogenous concentrations, the efficiency of membrane fusion was lower than that obtained in the presence of cytosol. However, a combination of EEA1 and fraction 27 (containing Rabaptin-5, Rabaptin-5 β , Rabex-5 and other proteins) resulted in full recovery of the fusion efficiency. Addition of recombinant Rabaptin-5/Rabex-5 complex also stimulated the fusion activity although less potently than fraction 27, suggesting that other unidentified proteins present in the fraction cooperate with the Rabaptin-5 complex. Fractions 36 and 39 could also stimulate the activity of EEA1, albeit less efficiently than fraction 27. These data argue strongly that different Rab5-interacting proteins can exert a regulatory function on EEA1-mediated early-endosome fusion.

As EEA1 alone significantly supported endosome fusion in the absence of cytosol, we investigated the role of this protein further. If EEA1 is a bona-fide core component of the docking and fusion machinery, and if other Rab5 effectors modulate the nucleotide binding of Rab5 (for example, the Rabaptin-5/Rabex-5 complex)

and the membrane recruitment and activity of EEA1, this regulation may be bypassed by increasing levels of EEA1. Figure 2c shows that 75 nM EEA1 (threefold concentration compared with cytosol) was able to drive fusion between early endosomes with similar efficiency to cytosol. Higher concentrations of EEA1 markedly enhanced fusion efficiency. Addition of fraction 27 enhanced this activity but fusion became insensitive to fraction 27 when excess EEA1 was supplied (Fig. 2c). The rate of fusion obtained with excess EEA1 was tenfold higher than the cytosol-driven reaction (not shown). These results indicate strongly that EEA1 is an integral component of the docking and fusion machinery.

Under native conditions of endosome fusion (that is, in the presence of cytosol) the recruitment of EEA1 to the endosome is dependent on Rab5 and phosphatidylinositol-3-phosphate (PtdIns(3)P)^{18,23}. Although GDP dissociation inhibitor (GDI) inhibited the cytosol-dependent fusion between early endosomes, it had no effect on fusion induced by 75 nM or 225 nM EEA1 (Fig. 3a). However, fusion under stimulatory conditions provided by fraction 27 demonstrates a partial GDI sensitivity. Therefore, affinity-purified EEA1 bypasses the requirement for Rab5 but the enhancement of its activity by fraction 27 is Rab5-dependent. In addition, the requirement for PtdIns(3)P on EEA1-mediated fusion was tested by pre-treating endosomes with wortmannin to inhibit phosphatidylinositol-3-OH kinase (PI(3)K)^{18,24}. Although unable to fuse in the presence of cytosol, wortmannin-treated endosomes fused efficiently in the presence of excess EEA1 (Fig. 3a). Therefore, as for Rab5, when EEA1 is supplied as the only cytosolic protein factor it can bypass the requirement for PtdIns(3)P. This indicates Rab5 and PtdIns(3)P as upstream regulators of EEA1.

Peripheral endosomal membrane proteins did not seem to be required (or rate-limiting) for the EEA1-mediated fusion as salt-washed endosomes fused with equal efficiency to the control endosomes. Under these conditions, NSF was only partially removed by the salt wash (not shown), consistent with its tight association with the membrane²¹. Therefore, there are presumably sufficient amounts of NSF to ensure priming of SNAREs. Altogether, these results indicate that EEA1 is a component of the minimal endocytic machinery for membrane docking and fusion.

To determine the hierarchical order of function between EEA1 and SNAREs, we examined the effect of various reagents known to block SNARE priming. First, we found that EEA1-mediated fusion was inhibited by 60% and 40% by ATP- γ S and *N*-ethylmaleimide (NEM) treatment, respectively (Fig. 3b). Similar inhibition was observed by ATP depletion (not shown). Second, we took advantage of a more specific inhibitor. We used a mutant of α -SNAP, α -SNAP(L294A), which, by inhibiting the stimulation of NSF ATPase activity, fails to prime SNAREs²⁵. When added to a fusion reaction conducted either with cytosol or in the presence of EEA1 with fraction 27, α -SNAP(L294A) caused a strong, dose-dependent inhibition of fusion. At high concentrations of α -SNAP(L294A) (20 μ M), fusion was completely abolished. The reaction was also inhibited to a similar extent when excess EEA1 alone was used. In contrast, wild-type α -SNAP did not inhibit endosome fusion. These data indicate that the NSF- and SNAP-dependent priming of SNAREs is at least partially required for EEA1-mediated fusion of early endosomes.

At which stage does EEA1 become important in the docking and fusion reaction? As the assay of homotypic endosome fusion cannot distinguish between docking and fusion, we developed a morphological docking assay. For this purpose, rhodamine-labelled transferrin was internalized in HeLa cells and endosomes were purified. These membranes were incubated at 37 °C for 25 min with various reagents (Fig. 4a) and subsequently observed under a fluorescence microscope. The combination of the fusion and fluorescence assays enabled us to determine the docked state of the membranes. Under experimental conditions that do not produce fusion, that is, in the absence of cytosol, only small scattered fluorescent structures were

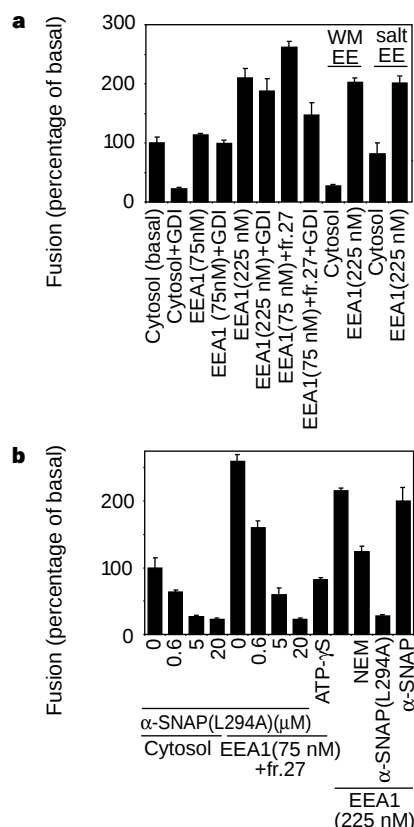


Figure 3 EEA1-mediated fusion can bypass Rab5 and PtdIns(3)P, but is inhibited by α -SNAP(L294A), ATP- γ S and NEM. **a**, Endosome fusion was performed in the presence of the indicated reagents. GDI was added to a final concentration of 18 μ M. Wortmannin treatment of the early endosomes (WM EE) was done by incubating post-nuclear supernatant (PNS) fraction with 1 μ M wortmannin for 20 min at room temperature. Subsequently, endosomes were isolated on a sucrose flotation gradient as with normal endosomes. Similarly, salt-washed endosomes (salt EE) were obtained by incubating PNS with 1 M final concentration of NaCl on ice for 30 min, followed by isolation of the endosomes on a sucrose flotation gradient. **b**, Fusion assay was done in the presence of cytosol (basal) or the indicated reagents. ATP- γ S was used at concentration 1 mM in the absence of an ATP regenerating system. The NEM concentration was 3 mM.

seen (not shown). Larger structures formed in the presence of cytosol, resulting from the basal docking and fusion activity. EEA1 alone markedly stimulated the enlargement of endosomes and its effect was enhanced by the concomitant addition of fraction 27. The same morphological effect was obtained in the presence of α -SNAP(L294A), despite the fact that under these conditions membrane fusion is inhibited. Therefore, these clustered structures must be due predominantly to docking.

Consistent with these results, ATP- γ S also had only a marginal effect on docking, although fusion was inhibited. We quantified the docked clusters and compared the values with the data from the fusion assay (Fig. 4b). Both α -SNAP(L294A) and ATP- γ S are thought to prevent SNARE priming^{1,25,26}. Here, they have no effect on EEA1-mediated endosome docking. Only at higher concentrations of α -SNAP(L294A) (20 μ M), beyond the levels required to prevent SNARE priming²⁵, did we observe an inhibition of endosome docking. Golgi membranes failed to dock in the presence of EEA1 (not shown), demonstrating that the effect of EEA1 is specific for endosomes. After the initial incubation in the presence of EEA1 and α -SNAP(L294A) to block fusion, membranes were diluted. No dispersion of the membranes was observed, which indicates that the membranes were tightly associated and resistant to dilution (not shown).

We have shown that Rab5 is at the centre of a complex regulatory

system involving multimeric interactions. Surprisingly, the Rab5 effector EEA1 alone can support fusion between early endosomes. We propose that EEA1 is a core component of the minimal machinery that docks and fuses endosomes. Our results suggest the following sequence of events. The activation of Rab5 on the endosome is ensured by the Rabaptin-5/Rabex-5 complex¹⁶. EEA1 recruitment and activity is regulated by active Rab5 and PtdIns(3)P¹⁸. In a minimal system, however, it is possible to bypass these regulatory mechanisms and proceed to a further downstream step that leads directly to docking and fusion. In fact, in the absence of cytosol and with EEA1 as the only soluble factor, endosome fusion becomes uncoupled from the Rab5- and PtdIns(3)P-dependent regulation. This observation has two implications. First, membrane factors other than Rab5 and PtdIns(3)P must participate in the recruitment and activation of EEA1 and, second, the activity of EEA1 may normally be counteracted by cytosolic inhibitory factors. Rab5 and its effectors would then be necessary to override this negative regulation. Again, in our minimal system this negative regulation is either lost or, alternatively, purified EEA1 may have become activated as a result of the interaction with Rab5 on the column. In addition, given the stimulatory and Rab5-dependent effect of the column-eluate fractions, other Rab5 effectors can modulate EEA1 recruitment and/or function.

By exploiting a combination of morphological and biochemical assays, we now propose a role for EEA1 in endosome membrane docking (referred to also as tethering^{8,27}). Although tethering molecules have been identified^{27,28}, this is the first direct evidence for such a role by a Rab effector. The function of EEA1 in docking, together with its restricted localization to early endosomes²⁹, implies that this molecule is crucial in the targeting of vesicles to this organelle. EEA1-mediated docking does not seem to require v- and t-SNARE priming as it was not affected by ATP- γ S and α -SNAP(L294A). The docking of endoplasmic reticulum vesicles to Golgi has also been shown to occur in the absence of SNARE priming through the tethering protein Usp1p⁸. These results are therefore incompatible with a primary role of SNAREs in vesicle targeting. We consider it more likely that the specificity of vesicle targeting results from the combinatorial interaction between different cytosolic and membrane proteins (Rab effectors, SNAREs and perhaps others).

Although not required for docking, SNARE priming seems to be necessary for fusion. It is not inconceivable, however, that EEA1 may also participate in the fusion event itself. Three lines of evidence support this hypothesis. First, overexpression of a truncated mutant of EEA1 arrests the membranes in a docked stage and inhibits endosome fusion *in vivo*¹⁸. Second, we could not distinguish kinetically between EEA1-mediated docking and fusion (data not shown), indicating that the two stages are tightly coupled in the early-endosome system. Third, EEA1-mediated fusion is only partially ATP- and NEM-sensitive. One interpretation is that EEA1 may, to a certain extent, fuse endosomes in a SNARE-independent fashion. However, the observation that fusion is strongly inhibited by α -SNAP(L294A) indicates that this mutant could transcend the inhibition of NSF ATPase activity, raising interesting possibilities for a new role of α -SNAP (and NSF) in endocytic membrane fusion. Finally, the multiplicity of putative Rab5 effectors encourages us to consider the role of Rab5 in regulating other aspects of endosome function beyond docking and fusion.

□

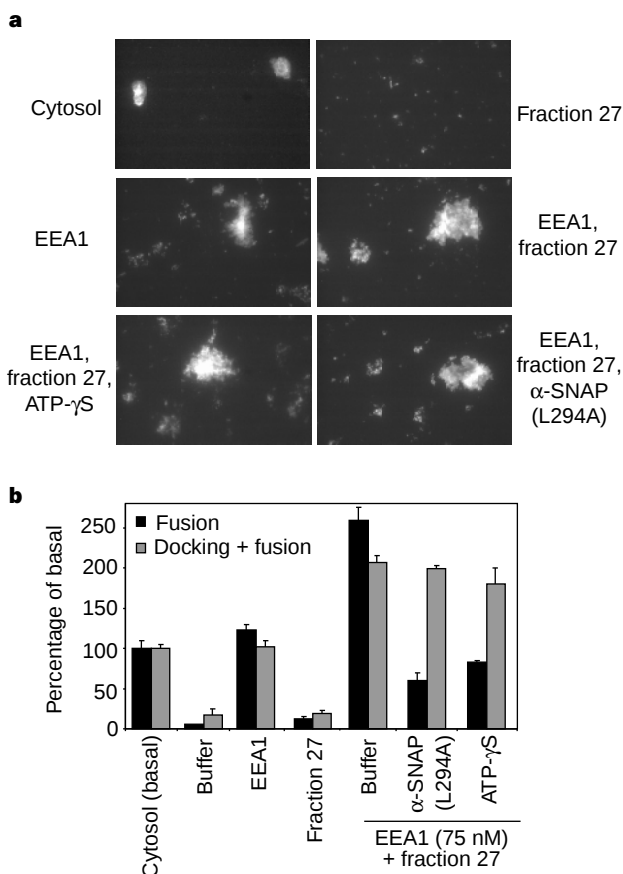


Figure 4 Endosome clustering requires EEA1 but not primed SNAREs. **a**, Visual docking of endosomes. Early endosomes, labelled with rhodamine-labelled transferrin, were incubated with cytosol or the indicated reagents; after 25 min the reaction was visualized with a Zeiss Axiophot fluorescence microscope. Concentrations used were: EEA1, 75 nM; fraction 27, 1.5 μ l per 10 μ l reaction; α -SNAP(L294A), 5 μ M; ATP- γ S, 1 mM. **b**, Comparison between fusion and docking plus fusion. Fusion (solid bars) is measured as a percentage of basal (in the presence of cytosol). Quantitation of docking (shaded bars) was done as described (see Methods). Reagents were used at the concentration described in **a**.

Methods

Purification of Rab5 effectors. Bacterial DH5 α cells (120 l) were grown to express GST-Rab5 and the protein was purified according to manufacturer instructions (Pharmacia). This procedure resulted in 1 g of GST-Rab5 bound to 20 ml of glutathione Sepharose 4B beads. This material was then incubated with nucleotide-exchange buffer (buffer 1) containing 20 mM HEPES, 100 mM

NaCl, 10 mM EDTA, 5 mM MgCl₂, 1 mM DTT, 1 mM GTP-γS, pH 7.5, for 90 min at room temperature under rotation. Afterwards, buffer 1 was removed and the GTP-γS form of GST-Rab5 was stabilized with buffer (buffer 2) containing 20 mM HEPES, 100 mM NaCl, 5 mM MgCl₂, 1 mM DTT, pH 7.5, in the presence of 1 mM GTP-γS for 20 min at room temperature under rotation. Beads were then incubated for 120 min at 4 °C with bovine brain cytosol obtained as follows: 14 bovine brains were homogenized in a blender with buffer 2 and the homogenate was centrifuged at 4,200g at 4 °C for 50 min. The resulting postnuclear supernatant was then centrifuged at 100,000g at 4 °C for 60 min. The high-speed supernatant was dialyzed against buffer 2 (without nucleotide) before incubation with the affinity column. After incubation with cytosol, beads were washed with ten column volumes of buffer 2 containing 10 μM GTP-γS, ten column volumes of buffer 2 containing 250 mM NaCl final concentration and 10 μM GTP-γS, and one column volume of 20 mM HEPES, 250 mM NaCl, 1 mM DTT, pH 7.5.

Bound proteins were eluted with 1.5 column volumes of a buffer containing 20 mM HEPES, 1.5 M NaCl, 20 mM EDTA, 1 mM DTT, 5 mM GDP, pH 7.5, incubated with the beads for 20 min at room temperature under rotation. EDTA was used at this step to remove Mg²⁺ from Rab5 and release the effectors from the column. As EEA1 is a Zn²⁺-binding protein and EDTA chelates Zn²⁺ (refs 23, 30), our subsequent assays were done in the presence of 1 mM ZnCl₂ filtered (0.22 μM) before its use. The GDP form of GST-Rab5 (1 ml) was made as described above for the GTP-γS form of GST-Rab5 with the following changes: all buffers contained GDP instead of GTP-γS, except for the elution buffer which contained GTP-γS (1 mM) instead of GDP. In the case of GST column (100 μl), there was no nucleotide in the buffers.

Fractionation of Rab5 effectors. The eluate (30 ml) from the 20-ml affinity column containing the mixture of Rab5-interacting proteins was first treated twice for 1 h at 4 °C with 1.5 ml glutathione Sepharose beads to remove GST-Rab5 which leaked from the affinity column during the elution step. The sample was then desalted using PD10 columns from Pharmacia in a buffer containing 20 mM HEPES, 150 mM NaCl, 1 mM DTT, pH 7.5, and diluted three times with 20 mM HEPES, 1 mM DTT, pH 7.5, resulting in a final volume of 135 ml. The diluted sample was loaded on a 1-ml MQ FPLC column (Pharmacia) and bound proteins were step-eluted with 20 mM HEPES, 1 M NaCl, 1 mM DTT, pH 7.5, in a total volume of 1 ml (concentration step). This eluate was fractionated on a 24-ml Superose 6 FPLC gel filtration column (Pharmacia). Fractions of 0.4 ml were collected, aliquoted and frozen at -80 °C.

Endosome docking assay. HeLa cells grown in suspension were harvested, washed with PBS and incubated for 5 min with 20 μg ml⁻¹ rhodamine-labelled transferrin at 37 °C for 5 min. Then endosomes were isolated¹⁵ and incubated in a 10 μl reaction for 25 min at 37 °C with reagents mentioned in Fig. 4. At the end of the reaction, samples were put on ice and visualized with a Zeiss Axiophot fluorescence microscope. Images were taken using a Cohu camera. Quantitation of docking was done by taking random pictures, followed by counting of the occupied squares (an indication of organelle area) on a reference grid.

Other preparations. GST-Rab5 and GST proteins were expressed in *Escherichia coli* using the P-GEX vector (Pharmacia). his-Rab GDI¹⁶, his-α-SNAP and his-α-SNAP(L294A)²⁵ were produced as described. The preparation of recombinant Rabaptin-5/Rabex-5 complex will be described elsewhere (R. Lippe and M. Zerial, manuscript in preparation). Early endosome fusion assay was done as described¹⁶.

Received 9 October 1998; accepted 8 January 1999.

1. Rothman, J. E. Mechanisms of intracellular protein transport. *Nature* **372**, 55–63 (1994).
2. Weber, T. et al. SNAREpins: Minimal machinery for membrane fusion. *Cell* **92**, 759–772 (1998).
3. Robinson, L. M. & Martin, T. F. J. Docking and fusion in neurosecretion. *Curr. Opin. Cell Biol.* **10**, 483–492 (1998).
4. von Mollard, G. E., Nothwehr, S. F. & Stevens, T. H. The yeast v-SNARE Vti1p mediates two vesicle transport pathways through interactions with the t-SNAREs Sed5p and Pep12p. *J. Cell Biol.* **137**, 1511–1524 (1997).
5. Holthuis, J. C., Nichols, B. J., Dhruvakumar, S. & Pelham, H. R. Two syntaxin homologues in the TGN/endosomal system of yeast. *EMBO J.* **17**, 113–126 (1998).
6. Hay, J. C. et al. Localization, dynamics, and protein interactions reveal distinct roles for ER and Golgi SNAREs. *J. Cell Biol.* **141**, 1489–1502 (1998).
7. Sapperstein, S. K., Lupashin, V. V., Schmitt, H. D. & Waters, M. G. Assembly of the ER to Golgi SNARE complex requires Uso1p. *J. Cell Biol.* **132**, 755–767 (1996).
8. Cao, X., Ballew, N. & Barlowe, C. Initial docking of ER-derived vesicles requires Uso1p and Ypt1p but is independent of SNARE proteins. *EMBO J.* **17**, 2156–2165 (1998).
9. VanRheenen, S. M., Cao, X., Lupashin, V. V., Barlowe, C. & Waters, G. M. Sec35p, a novel peripheral membrane protein, is required for ER to Golgi vesicle docking. *J. Cell Biol.* **141**, 1107–1119 (1998).

10. Novick, P. & Zerial, M. The diversity of Rab proteins in vesicle transport. *Curr. Opin. Cell Biol.* **9**, 496–504 (1997).
11. Sogaard, M. et al. A Rab protein is required for the assembly of SNARE complexes in the docking of transport vesicles. *Cell* **78**, 937–948 (1994).
12. Lian, J. P., Stone, S., Jiang, Y., Lyons, P. & Ferro-Novick, S. Ypt1p implicated in v-SNARE activation. *Nature* **372**, 698–701 (1994).
13. Mayer, A. & Wickner, W. Docking of yeast vacuoles is catalyzed by the Ras-like GTPase Ypt7p after symmetric priming by Sec18p (NSF). *J. Cell Biol.* **136**, 307–317 (1997).
14. Lupashin, V. V. & Waters, M. G. t-SNARE activation through transient interaction with a rab-like guanosine triphosphatase. *Science* **276**, 1255–1258 (1997).
15. Gorvel, J.-P., Chavrier, P., Zerial, M. & Gruenberg, J. Rab5 controls early endosome fusion in vitro. *Cell* **64**, 915–925 (1991).
16. Horiuchi, H. et al. A novel Rab5 GDP/GTP exchange factor complexed to Rabaptin-5 links nucleotide exchange to effector recruitment and function. *Cell* **90**, 1149–1159 (1997).
17. Gournier, H., Stenmark, H., Rybin, V., Lippe, R. & Zerial, M. Two distinct effectors of the small GTPase Rab5 cooperate in endocytic membrane fusion. *EMBO J.* **17**, 1930–1940 (1998).
18. Simonsen, A. et al. EEA1 links phosphatidylinositol 3-kinase function to Rab5 regulation of endosome fusion. *Nature* **394**, 494–498 (1998).
19. Emans, N. et al. Annexin II is a major component of fusogenic endosomal vesicles. *J. Cell Biol.* **120**, 1357–1369 (1993).
20. Colombo, M. L., Beron, W. & Stahl, P. D. Calmodulin regulates endosome fusion. *J. Biol. Chem.* **272**, 7707–7712 (1997).
21. Rodriguez, L., Stirling, C. J. & Woodman, P. G. Multiple N-ethylmaleimide-sensitive components are required for endosomal vesicle fusion. *Mol. Biol. Cell* **5**, 773–783 (1994).
22. Colombo, M. L., Taddese, M., Whiteheart, S. W. & Stahl, P. D. A possible predocking attachment site for N-ethylmaleimide-sensitive fusion protein. Insights from in vitro endosome fusion. *J. Biol. Chem.* **271**, 18810–18816 (1996).
23. Wiedemann, C. & Cockcroft, S. Vesicular transport: sticky fingers grab a lipid. *Nature* **394**, 426–427 (1998).
24. Patki, V. et al. Identification of an early endosomal protein regulated by phosphatidylinositol 3-kinase. *Proc. Natl. Acad. Sci. USA* **94**, 7326–7330 (1997).
25. Barnard, R. J. O., Morgan, A. & Burgoyne, R. D. Stimulation of NSF ATPase activity by alpha-SNAP is required for SNARE complex disassembly and exocytosis. *J. Cell Biol.* **139**, 875–883 (1997).
26. Mayer, A., Wickner, W. & Haas, A. Sec18p (NSF)-driven release of Sec17p (α-SNAP) can precede docking and fusion of yeast vacuoles. *Cell* **85**, 83–94 (1996).
27. Sonnichsen, B. et al. A role for giantin in docking COPI vesicles to Golgi membranes. *J. Cell Biol.* **140**, 1013–1021 (1998).
28. Barlowe, C. Coupled ER to golgi transport reconstituted with purified cytosolic proteins. *J. Cell Biol.* **139**, 1097–1108 (1997).
29. Mu, F. et al. EEA1, an early endosome-associated protein. *J. Biol. Chem.* **270**, 13503–13511 (1995).
30. Stenmark, H., Aasland, R., Toh, B. H. & D'Arrigo, A. Endosomal localization of the autoantigen EEA1 is mediated by zinc-binding FYVE finger. *J. Biol. Chem.* **271**, 24048–24054 (1996).

Acknowledgements. We thank R. Lippe for the supply of recombinant Rabaptin-5/Rabex-5 complex; H. Stenmark for the GST-Rab5 cDNA construct; J. Rothman for pQE9 α-SNAP; A. Giner for technical assistance; V. Rybin for helpful discussions on technical points; E. Nielsen and B. Sonnichsen for the supply of rhodamine-labelled early endosomes; P. Scheiffele, B. Sonnichsen and members of the laboratory for discussions and critical reading of the manuscript; and K. Ashman and M. Wilm form providing mass spectroscopy data. S.C. is supported by an EU TMR fellowship. H.M.M. is recipient of an Alexander von Humboldt Stiftung. This work was supported by the Max Planck Gesellschaft, and by grants from the Human Frontier Science Program, EU TMR and Biomed (to M.Z.). This work is dedicated to the memory of Thomas Kreis.

Correspondence and requests for materials should be addressed to M.Z. (e-mail: Zerial@embl-Heidelberg.de).

Photosynthetic control of chloroplast gene expression

Thomas Pfannschmidt*, Anders Nilsson & John F. Allen

Plant Cell Biology, Lund University, Box 7007, S-220 07 Lund, Sweden

Redox chemistry—the transfer of electrons or hydrogen atoms—is central to energy conversion in respiration and photosynthesis. In photosynthesis in chloroplasts, two separate, light-driven reactions, termed photosystem I and photosystem II, are connected in series by a chain of electron carriers^{1–3}. The redox state of one connecting electron carrier, plastoquinone, governs the distribution of absorbed light energy between photosystems I and II by controlling the phosphorylation of a mobile, light-harvesting, pigment–protein complex^{4,5}. Here we show that the redox state of plastoquinone also controls the rate of transcription of genes encoding reaction-centre apoproteins of photosystem I and photosystem II. As a result of this control, the stoichiometry between the two photosystems changes in a way that counteracts the inefficiency produced when either photosystem limits the rate of the other. In eukaryotes, these reaction-centre proteins are encoded universally within the chloroplast. Photosynthetic control

* Present address: Plant Cell Physiology and Molecular Biology, University of Bochum, Universitätsstrasse 150, 44780 Bochum, Germany.

of chloroplast gene expression indicates an evolutionary explanation for this rule: the redox signal-transduction pathway can be short, the response rapid, and the control direct.

Photosynthesis and plant growth can be supported by light with a wavelength composition that favours either photosystem II ('light 2') or photosystem I ('light 1')^{2,3}. Figure 1 shows the effects of such photosystem-specific light on chloroplast composition and chloroplast gene expression. Figure 1a, b shows the relative transcriptional rates of specific chloroplast genes in 7-day-old mustard plants grown in light 1, in light 2, and after changes between the two regimes. Plants described as 'light 2 → 1' were grown for 5 days under photosystem-II light, before being transferred to photosystem-I light for 2 days; 'light 1' plants were grown under light 1 for 7 days; white-light-grown plants were used as the control; 'light 2' plants were grown under light 2 for 7 days; and 'light 1 → 2' plants were grown for 5 days under photosystem-I light, and then transferred to photosystem-II light for 2 days. Cotyledons of seedlings were collected, and chloroplasts were isolated and used in chloroplast run-on transcription experiments. This method leads to radioactive pulse-labelling of messenger RNA

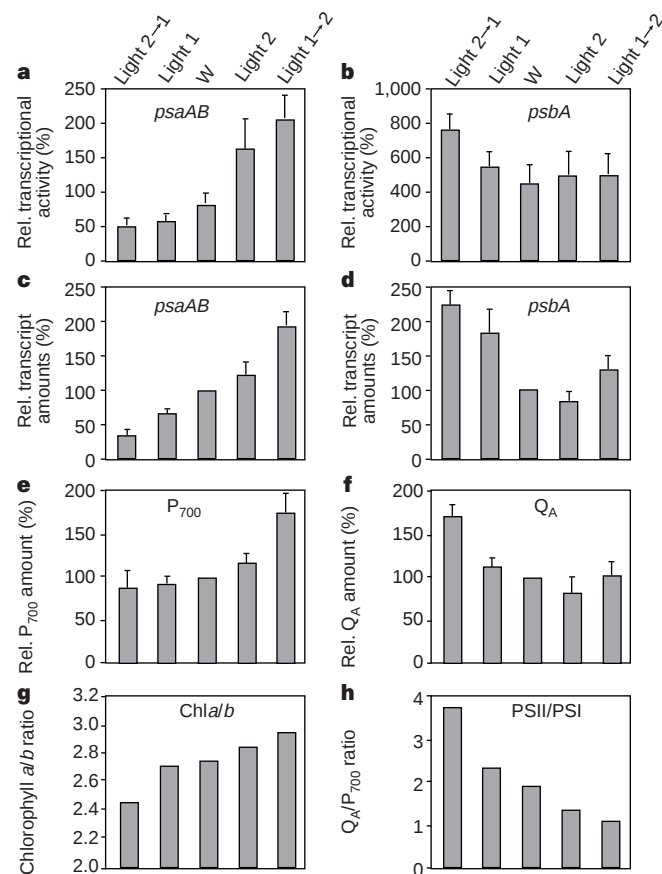


Figure 1 Reaction-centre stoichiometry and transcription. Transcriptional rate (a, b) and mRNA abundance (c, d) for the chloroplast reaction-centre genes *psaAB* (photosystem I) and *psbA* (photosystem II), together with quantities of the functional prosthetic groups, P_{700} (e) and Q_A (f), to which their respective apoprotein gene products bind, and chlorophyll *a/b* (g) and photosystem II/I (PS II/PS I) ratios (h). Gene expression and functional consequences for photosystem I are arranged in the left-hand column (a, c, e, g); those for photosystem II are arranged in the right-hand column (b, d, f, h). The PS II/PS I ratios in (h) are the ratios of Q_A to P_{700} from (f) and (e). Values in a and b are means of between four and five independent experiments. Values in c-f are means of three independent experiments and are expressed as percentages of the value of the white-light control (W).

from a gene that was actually being transcribed at the time of chloroplast isolation. The transcriptional rate of the genes for the two major reaction-centre subunits of photosystem I, *psaAB*, shows an incremental increase from left to right (Fig. 1a), as photosystem I changes from being light-saturated to being light-limited. The transcriptional rate for the *psbA* gene, which encodes the D_1 reaction-centre protein of photosystem II, shows broadly the reverse pattern (Fig. 1b): the transcriptional rate of *psbA* is highest when photosystem II is rate-limiting ('light 2 → 1'). The RNA quantities of *psaAB* (Fig. 1c) and, to a lesser extent, *psbA* (Fig. 1d), follow changes in their respective transcriptional rates (Fig. 1a, b).

Figure 1e, f shows that the quantity of functional reaction centres follows the same pattern of response to illumination as the rate of transcription (Fig. 1a, b) and as transcript pool size (Fig. 1c, d). P_{700} is the specialized chlorophyll *a* that is bound to the *psaAB* gene products. It undergoes primary charge separation in photosystem I (ref. 6). P_{700} (Fig. 1e) is least abundant when the light supplied favours its activity (light 2 → 1), and most abundant when it does not (light 1 → 2). Q_A is the species of plastoquinone that is bound to the *psbA* gene product. It acts as a secondary electron acceptor in photosystem II (ref. 7). Q_A (Fig. 1f) is most abundant where illumination favours photosystem I (light 2 → 1), and least abundant where it favours photosystem II (light 2 and light 1 → 2). The chlorophyll *a/b* ratio (Fig. 1g) follows the same pattern as P_{700} and as the rate of transcription of *psaAB*. The ratio of photosystem II to I, measured as the molar ratio of Q_A to P_{700} (Fig. 1h), follows the same pattern as the absolute quantity of Q_A and as the rate of transcription of *psbA*. These results indicate that when either photosystem becomes rate-limiting to photosynthesis, transcription of genes for its reaction-centre proteins is induced. Genes for reaction-centre proteins of the other photosystem, which has surplus photochemical capacity, are simultaneously repressed.

Figure 2 shows that the light-induced changes in rate of transcription and quantity of transcript are remarkably fast. For both *psaAB* (Fig. 2a) and *psbA* (Fig. 2b) the change in rate of transcription, whether an increase or decrease, takes place over a timescale of minutes, and is followed by a corresponding change in quantity of

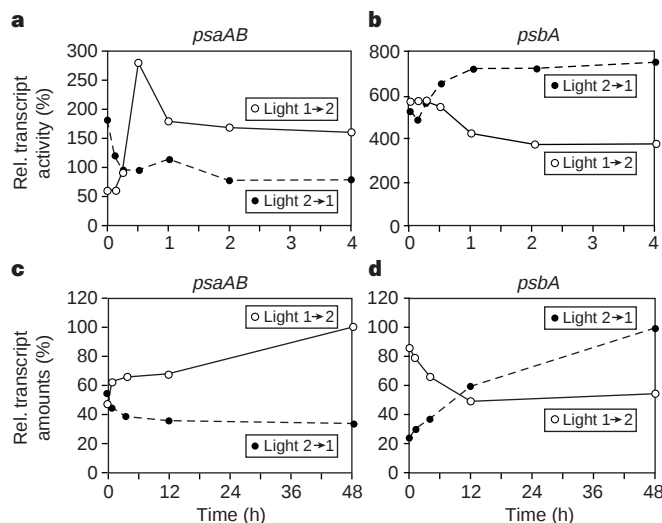


Figure 2 Complementary changes in transcription. Changes in transcriptional rate (a, b) and mRNA abundance (c, d) for the chloroplast reaction-centre genes *psaAB* (photosystem I; a, c) and *psbA* (photosystem II; b, d), as a function of time after changes in light regime between illumination conditions favouring photosystem I and photosystem II. Each change in illumination took place after 5 days' growth in either light 1 or light 2. Changes in rate of *psaAB* transcription (a) are detectable within 15 min, during which time fluorescence-emission changes that report on post-translational events are still taking place (results not shown). Each value is the mean of three independent experiments.

mRNA (Fig. 2c, d). We conclude that changes in the rate of transcription (Fig. 2a, b) arise from light-induced perturbation of the redox state of the plastoquinone pool. The altered rates of transcription result, in turn, in altered mRNA quantities (Fig. 2c, d) and, ultimately, in the altered stoichiometries of the two photosystems of the chloroplast thylakoid membrane (Fig. 1e–h).

The inhibitor 3-(3',4'-dichlorophenyl)-1,1'-dimethyl urea (DCMU) inhibits electron flow from photosystem II into the plastoquinone pool⁸. A complementary electron-transport inhibitor is 2,5-dibromo-3-methyl-6-isopropyl-p-benzoquinone (DBMIB), which inhibits oxidation of plastoquinol by photosystem I (ref. 8). Figure 3 shows that the rate of *psaAB* transcription in isolated chloroplasts is decreased in the presence of DCMU or by light 1, and increased in the presence of DBMIB or by light 2. These effects on transcriptionally and photosynthetically active chloroplasts isolated *in vitro* are precisely as predicted if plastoquinone redox control accounts for the responses observed *in vivo* (Figs 1, 2), and demonstrate that the control does not depend upon nuclear or cytosolic components. Taken together with the observation that the *psaAB* transcriptional rate *in vivo* is higher in light 2 than in light 1 (Fig. 1), we conclude that *psaAB* transcription is promoted when plastoquinone is reduced, and retarded when it is oxidized. In contrast, the rate of transcription of *psbA* (Fig. 3) does not respond to light or inhibitors *in vitro* under our experimental conditions, and seems to be a function of plastoquinone redox state only *in vivo* (Figs 1, 2).

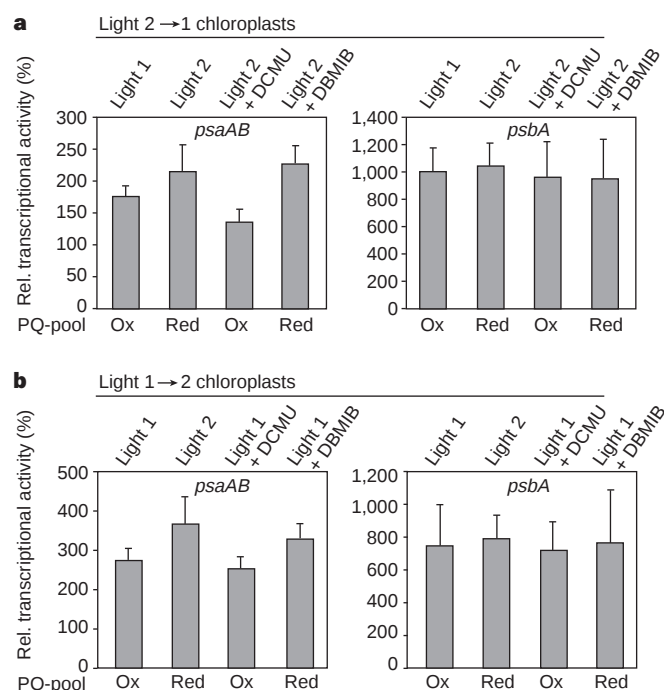


Figure 3 Transcription of reaction-centre genes *psaAB* (photosystem I) and *psbA* (photosystem II) in chloroplasts isolated from 7-day-old light 2 → 1 and light 1 → 2 plants. For *psaAB*, transcription is increased in light 2 relative to that in light 1, consistent with the effects observed *in vivo* (Fig. 2). Light 2 is known to produce a reduced plastoquinone pool. The inhibitor DBMIB also results in net reduction of plastoquinone, by suppressing oxidation of plastoquinol by photosystem I: the effect of DBMIB in increasing transcription of *psaAB* is consistent with redox control at the level of plastoquinone. The inhibitor DCMU causes oxidation of the plastoquinone pool by suppressing its reduction by photosystem II: the effect of DCMU in decreasing *psaAB* transcription supports the conclusion that *psaAB* is transcribed when plastoquinone is reduced, but not when it is oxidized. Redox effects on *psbA* transcription, in contrast, are absent under the conditions of this experiment. Error bars represent variations of four independent experiments.

Although various redox effects have been described for translation of chloroplast *psbA*⁹, transcription of nuclear genes^{10–12} and *de novo* protein synthesis in isolated chloroplasts¹³, our results show a new, direct regulatory coupling between the redox state of plastoquinone and transcription of specific chloroplast genes. Figure 4 describes how this coupling supports a novel mechanism for physiological and developmental adjustment of photosystem stoichiometry¹⁴.

There is an additional, evolutionary implication of these results. The ancestor of eukaryotic cells acquired many genes upon its merger^{15–17} with the prokaryotic, eubacterial ancestors of chloroplasts¹⁸ and mitochondria¹⁹. Most of the genes subsequently

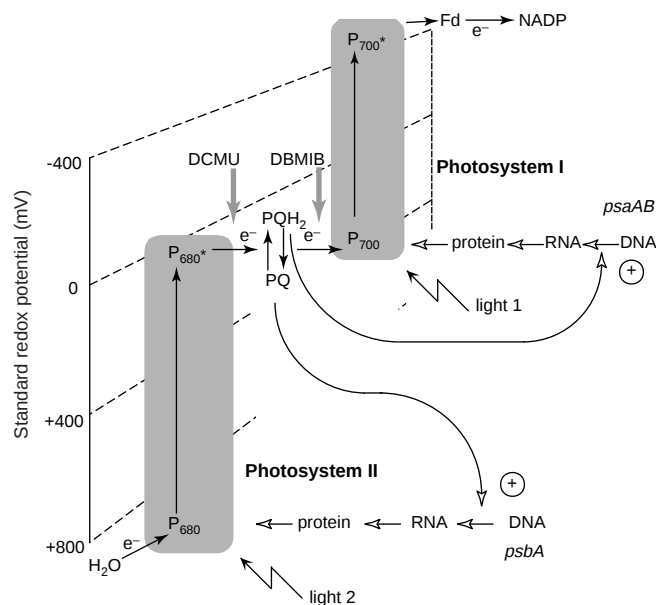


Figure 4 A working hypothesis for photosynthetic control of chloroplast gene expression by redox regulation of transcription. The Hill and Bendall 'Z-scheme' for non-cyclic electron transport in photosynthesis¹ is presented with standard redox potential of electron carriers plotted on the y-axis, and with the sequence of electron transfers plotted on a projected z-axis, at 30° to the horizontal. The x-axis itself schematically represents gene expression. Electron flow from H₂O to NADP⁺ requires two photosystems. The chemically reduced form of plastoquinone, plastoquinol (PQH₂), is produced by electron flow into the plastoquinone pool from photosystem II, and consumed by outward electron flow to photosystem I. Experimentally, electron flow from photosystem II to plastoquinone can be selected with light 2 and inhibited by addition of DCMU, whereas electron flow from plastoquinol to photosystem I can be selected with light 1 and inhibited by addition of DBMIB. *psaAB*, which encodes reaction-centre apoproteins of photosystem I, is induced when plastoquinone is reduced, and repressed when it is oxidized (PQ). *psbA*, encoding the D₁ reaction-centre protein of photosystem II, is induced when plastoquinone is oxidized and repressed when plastoquinone is reduced. The plastoquinone pool therefore exerts regulatory control over synthesis of the two photosystems in such a way as to balance its own reduction by photosystem II with its own oxidation by photosystem I. This mechanism of redox control determines the stoichiometry of photosystem I to photosystem II (Fig. 1), and adjusts this stoichiometry in response to any change in illumination that favours one or other of the two photosystems. Photosystem stoichiometry adjustment is thus a transcriptional counterpart to the more rapid, post-translational adjustment of the relative light-harvesting function of the two photosystems^{4,5}, and serves a similar function in maintaining redox homeostasis and photosynthetic efficiency. Although redox signals may also, indirectly, control nuclear gene expression^{10–12}, location of reaction-centre genes within the chloroplast itself permits transcriptional responses within minutes of perturbation of the redox state of the plastoquinone pool (Fig. 2).

retained have now been removed to the cell nucleus, but a small and relatively constant sub-set of genes has remained *in situ*, within the organelle. So why do chloroplasts and mitochondria retain any genes at all? One proposal is that chloroplast and mitochondrial genetic systems are required to permit direct redox regulation of gene expression^{20,21}. Our results are consistent with this hypothesis, and identify plastoquinone or a near neighbour as the site of a rapid and direct redox control of chloroplast transcription. The transcriptional responses shown here are specific and are many times more rapid (Fig. 2a) than comparable effects on nuclear genes^{10–12}. These kinetics are reminiscent of transcriptional control in prokaryotic systems. Such rapid and direct regulatory coupling may depend upon the genes concerned being present in the same intracellular compartment as the electron-transport chain that regulates their expression, and upon the persistence there of prokaryotically derived, redox signal-transduction pathways to provide the means of control. □

Methods

Plant material and illumination conditions. Mustard seedlings (*Sinapis alba* L.) were grown on vermiculite at 22 °C under light 1, light 2, white light (7 days) or light 1/light 2, light 2/light 1 (5/2 days) in a 16:8 h light–dark cycle at 30–35 $\mu\text{E m}^{-2} \text{s}^{-1}$. Light 1 was provided by 40-W incandescent light bulbs and filtered through a red filter giving 50% transmittance at 650 nm (Lee Filters, 027 Medium Red); light 2 was provided by 30-W cool-white fluorescent strip lamps and filtered through an orange-yellow filter giving 50% transmittance at 560 nm (Strand Lightning Filters, 405 Orange). For white light, a combination of incandescent bulbs and cool-white fluorescent strips was used without additional spectral filtering. Measurements of the modulated chlorophyll-fluorescence emission confirmed the selective effects of light 1 and light 2 on photosystems I and II, respectively²²: light 1 oxidized plastoquinol and induced the transition to light-state 1, whereas light 2 reduced plastoquinone and induced the transition to light-state 2 (refs 3–5).

Chloroplast isolation and spectroscopy. Chloroplasts and thylakoid membranes were prepared by a method based on ref. 23. Concentrations of Q_A and P_{700} were determined spectroscopically according to ref. 24 using the extinction coefficients of 11 $\text{mM}^{-1} \text{cm}^{-1}$ for Q_A (ref. 7) and 64 $\text{mM}^{-1} \text{cm}^{-1}$ for P_{700} (ref. 6). Chlorophyll concentrations of plant cell subfractions and of total plant material were determined spectroscopically after extraction in 80% (v/v) buffered acetone using the extinction coefficients of ref. 25.

Chloroplast run-on transcription. Run-on assays were performed essentially as described²⁶. Between 2.5×10^7 and 3×10^7 plastids were used in the standard assay. DNA probes representing mustard chloroplast genes for detection of labelled transcripts in run-on transcription experiments were *psaAB* (pSA224-EBH1.9)²⁷, *psbA* (pSA452a)²⁸ and *rrn16* (pBSH895)²⁹. Gene-specific expression rates were quantified in phosphorimaging analyses. Values were normalized to the value for *rrn16*, which was found not to be influenced by the light sources, and are means from between four and five independent experiments. In inhibitor and light-effect experiments, plastids were pre-incubated with 0.5 μM DCMU or 0.2 μM DBMIB in the presence of 20 mM $\text{Na}_4\text{P}_2\text{O}_7$ and 10 mM NaHCO_3 and exposed for 30 min to light 1 or light 2. They were then subjected to the standard procedure for run-on transcriptional assay, as described above.

Northern analysis. Isolation of total RNA and northern analyses were done using standard protocols³⁰. Hybridization probes for *psaAB* and *psbA* gene

transcripts were generated by *in vitro* transcription with T3 and T7 RNA polymerases³⁰ of plasmids pSA224-EBH1.9 (ref. 27) and pSA452a (ref. 28).

Received 12 November 1998; accepted 7 January 1999.

- Hill, R. & Bendall, E. Function of the two cytochrome components in chloroplasts, a working hypothesis. *Nature* **186**, 136–137 (1960).
- Duysens, L. N. M. & Ames, J. Function and identification of two photochemical systems in photosynthesis. *Biochim. Biophys. Acta* **64**, 243–260 (1962).
- Myers, J. Enhancement studies in photosynthesis. *Annu. Rev. Plant Physiol.* **22**, 289–312 (1971).
- Allen, J. F., Bennett, J., Steinback, K. E. & Arntzen, C. J. Chloroplast protein phosphorylation couples plastoquinone redox state to distribution of excitation energy between photosystems. *Nature* **291**, 25–29 (1981).
- Allen, J. F. Protein phosphorylation in regulation of photosynthesis. *Biochim. Biophys. Acta* **1098**, 275–335 (1992).
- Hiyama, T. & Ke, B. Difference spectra and extinction coefficients of P_{700} . *Biochem. Biophys. Acta* **267**, 160–171 (1972).
- van Gorkom, H. J. Identification of the reduced primary electron acceptor of photosystem II as a bound semiquinone anion. *Biochim. Biophys. Acta* **347**, 439–442 (1974).
- Trebst, A. Inhibitors in electron flow: tools for the functional and structural localization of carriers and energy conservation sites. *Methods Enzymol.* **69**, 675–715 (1980).
- Danon, A. & Mayfield, S. P. Light-regulated translation of chloroplast messenger RNAs through redox potential. *Science* **266**, 1717–1719 (1994).
- Escoubas, J.-M., Loumas, M., LaRoche, J. & Falkowski, P. G. Light intensity regulation of *cab* gene transcription is signaled by the redox state of the plastoquinone pool. *Proc. Natl Acad. Sci. USA* **92**, 10237–10241 (1995).
- Maxwell, D. P., Laudenbach, D. E. & Huner, N. P. A. Redox regulation of light-harvesting complex II and *cab* mRNA abundance in *Dunaliella salina*. *Plant Physiol.* **109**, 787–795 (1995).
- Karpinski, S., Escobar, C., Karpinski, B., Creissen, G. & Mullineaux, P. M. Photosynthetic electron transport regulates the expression of cytosolic ascorbate peroxidase genes in *Arabidopsis* during excess light stress. *Plant Cell* **9**, 627–640 (1997).
- Allen, C. A., Håkansson, G. & Allen, J. F. Redox conditions specify the proteins synthesised by isolated chloroplasts and mitochondria. *Redox Report* **1**, 119–123 (1995).
- Fujita, Y., Murakami, A. & Ohki, K. Regulation of photosystem composition in the cyanobacterial photosynthetic system: the regulation occurs in response to the redox state of the electron pool located between the two photosystems. *Plant Cell Physiol.* **28**, 283–292 (1987).
- Whately, J. M., John, P. & Whately, F. R. From extracellular to intracellular: the establishment of mitochondrial and chloroplasts. *Proc. R. Soc. Lond. B* **204**, 165–187 (1979).
- Cavalier-Smith, T. The simultaneous symbiotic origin of mitochondria, chloroplasts, and microbodies. *Ann. NY Acad. Sci.* **503**, 55–71 (1987).
- Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
- Ellis, R. J. The nuclear domination of chloroplast development. *Sci. Prog. Oxf.* **69**, 129–142 (1984).
- Attardi, G. & Schatz, G. Biogenesis of mitochondria. *Annu. Rev. Cell Biol.* **4**, 289–333 (1988).
- Allen, J. F. Control of gene expression by redox potential and the requirement for chloroplast and mitochondrial genomes. *J. Theor. Biol.* **165**, 609–631 (1993).
- Allen, J. F. & Raven, J. A. Free-radical-induced mutation vs redox regulation: costs and benefits of genes in organelles. *J. Mol. Evol.* **42**, 482–492 (1996).
- Allen, J. F., Mullineaux, C. W., Sanders, C. E. & Melis, A. State transitions, photosystem stoichiometry adjustment and non-photochemical quenching in cyanobacterial cells acclimated to light absorbed by photosystem I or photosystem II. *Photosynth. Res.* **22**, 157–166 (1989).
- Walker, D. A. Chloroplast (and grana): aqueous (including high carbon fixation ability). *Methods Enzymol.* **23**, 211–220 (1971).
- Melis, A. & Brown, J. S. Stoichiometry of system I and system II reaction centres and of plastoquinone in different photosynthetic membranes. *Proc. Natl Acad. Sci. USA* **77**, 4712–4716 (1980).
- Porra, R. J., Thompson, W. A. & Kriedemann, P. E. Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls *a* and *b* extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochim. Biophys. Acta* **975**, 384–394 (1989).
- Mullet, J. E. & Klein, R. R. Transcription and RNA stability are important determinants of higher plant chloroplast RNA levels. *EMBO J.* **6**, 1571–1579 (1987).
- Dietrich, G., Detschey, S., Neuhaus, H. & Link, G. Temporal and light control of plastid transcript levels for proteins involved in photosynthesis during mustard (*Sinapis alba* L.) seedling development. *Planta* **172**, 393–399 (1987).
- Link, G. Cloning and mapping of the chloroplast DNA sequences for two messenger RNAs from mustard (*Sinapis alba* L.). *Nucleic Acids Res.* **9**, 3681–3694 (1981).
- Pfannschmidt, T. & Link, G. The A and B forms of plastid DNA-dependent RNA polymerase from mustard (*Sinapis alba* L.) transcribe the same genes in a different developmental context. *Mol. Gen. Genet.* **257**, 35–44 (1997).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1989).

Acknowledgements. We thank G. Link for discussions and for providing mustard seeds and DNA probes, and A. Tullberg for assistance with the experiments on isolated chloroplasts. This work was supported by the Swedish Natural Sciences Research Council and the Swedish Council for Co-ordination and Planning of Research. T.P. was the recipient of a postdoctoral fellowship of the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to J.E.A. (e-mail: john.allen@plantcell.lu.se).

Soft focus

Software is the peg for this week's selection of recent introductions. Packages for application development, and updates to software packaged with lab instruments are included. And turn to nature.com for a new web-only Software Review.

Reference Manager 9.0

From RIS www.risinc.com
PC/Windows-friendly bibliographical software.

Historically, bibliographic software for the higher education market has often been Mac based. Reference Manager 9.0, from Research Information Systems, is aimed at the PC/Windows sector of the market. The Windows interface includes spell-checker and drag-and-drop tools, one-step and customized bibliography generation.

Reader Service No. 100

TurboSEC

From Perkin Elmer www.perkin-elmer.com
Generate molecular weight distribution analysis reports

PE Nelson's TurboSEC software is an application for generating molecular weight distribution analysis reports from Turbochrom Workstation and Turbochrom Client/Server Chromatography Data Systems. The application runs either as a user program from within Turbochrom software to facilitate automated results processing and reporting, or as a stand-alone post-analysis program for individual data file analysis. TurboSEC software is a Windows GUI based program running under Windows NT 4.0 or Windows 95 as a native 32-bit application.

Reader Service No. 101

Magnet 6.0

From Infolytica www.infolytica.com
A finite-element electromagnetic simulation package

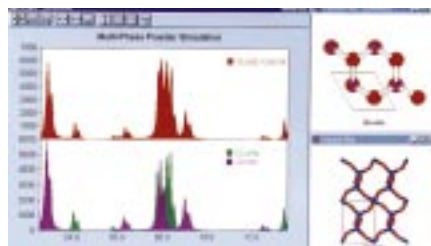
The latest release of MagNet 6.0 for PC is, say Infolytica, the product of a total re-write of the software. The new version features rapid simple model building and powerful manipulation, visualization, and animation in a single module. PC scripting facilities are included, using ActiveX compliant facilities (like Visual Basic or Java) for automatic model building. The software runs on PC586 or above, under Windows 95/98 or NT4.0.

Reader Service No. 102

Crystallographica

From Charles Supper www.charles-supper.com
Crystallography software taking Windows as a starting point

This new software utility is claimed to combine the ease-of-use of a spreadsheet with a



All-rounder: Crystallographica on screen.

fully integrated Pascal interpreter, extended with hundreds of 'building block' crystallographic routines. Crystallographica features a Pascal interpreter that lets users manipulate the variables in a powder simulation and observe the changes. Crystallographica performs X-ray and neutron powder pattern simulation and doubles as a teaching aid. This software utilizes Windows user interfaces and provides a viable alternative to one-off programs.

Reader Service No. 103

Automation

Microlink 552

From Biodata www.microlink.co.uk
A single PC can monitor a raft of transducers

Biodata's Microlink 552 package lets a laboratory PC monitor pH, temperature, strain, load, voltage, resistance, current and pressure sensors. Comprising Windmill data acquisition software for Windows, and a plug-in board, the system can handle a mix of up to 32 transducers. For more transducers, add more boards. Options that can be selected from software, include linearization for B, E, J, K, N, R, S and T type thermocouples. The Windmill data acquisition software is a modular suite providing, for example, data logging and charting. Windmill also works with other manufacturers' hardware and can export data to other applications, such as spreadsheets, in real-time.

Reader Service No. 104

ORCA Robotic Arm

From Beckman Coulter
www.beckmancoulter.com

Software upgrades for a familiar robot

Recent software enhancements for the ORCA (Optimized Robot for Chemical Analysis) Robotic Arm include a new 'Easy Calibration' feature, which brings automation to the calibration procedure. New ORCA NT 1.5 software features a wizard that guides the user through the Easy Calibration function. Built-in features make



Soft touch: calibration by wizards.

ORCA a frequent choice for automation for drug discovery, sample preparation, quality control and assurance, and clinical testing.

Reader Service No. 105

Mathematics

Mathwise

From NAG www.nag.co.uk
Learning maths — for Mac and Wintel

Mathwise, from the Numerical Algorithms Group, is a computer-based learning package for undergraduates in engineering and the sciences. Based on the SEFI (European Society for Engineering Education) syllabus, Mathwise sets out to provide an integrated learning environment for the teaching of mathematics. Mathwise consists of learning modules, each typically comprising 5 hours of lecture material. The program runs in both Windows 95 and on Macs.

Reader Service No. 106

TrueAllele

From Cybergenetics Inc.
www.cybergeneitics-inc.com

An evaluation CD-ROM to sample genotyping evaluation

Now available from Cybergenetics is an evaluation CD-ROM for its fully automated genetic analysis software. Mac, UNIX and PC platforms are catered for. TrueAllele genetic analysis software is marketed as a way around the data scoring bottleneck for microsatellite genetic data. The program automates data scoring, so can potentially accelerate gene discovery. Lane tracking and allele calling are automated on gel image data from Perkin Elmer/Applied

Biosystems and Pharmacia/ALF automated fluorescent DNA sequencers.

Reader Service No. 107

Mathcad 8 Treasury

From Adept Scientific www.adeptscience.co.uk

An interactive CD introduces calculating software

The Mathcad 8 Treasury is an interactive guide to the technical calculation software Mathcad 8 Professional, released in September of last year. The Treasury CD guides users through tips and techniques on how to apply Mathcad to their work and provides explanations of the mathematics behind each Mathcad 8 feature. Background information is given for each of the algorithms used in the program. The most recent version of Mathcad 8 Professional introduced extended mathematical functionality, enhanced document preparation facilities and improved the presentation capabilities. The program runs 3D graphics and now includes security features such as the ability to hide and password-protect sensitive or confidential information within a worksheet without



Summing up: background on Mathcad's maths.

affecting any of the calculations.

Reader Service No. 108

Microscopy

SPOT

From Diagnostic Instruments Inc

www.diaginc.com

Capturing microscopic images

Version 2.1.2 software for the Diagnostic Instruments SPOT cooled digital cameras is

designed specifically for microscopes. New features allow the user to capture, enhance, archive, print, annotate, and measure acquired images. The user can create image set-ups for many microscopy applications, and editing and annotation features may be applied to the image.

Reader Service No. 109

IPLab for Windows

From Scanalytics Inc www.scanalytics.com

Image acquisition software

Scanalytics' IPLab scientific image processing software for Windows is now into version 2.3. Developed for Windows 95/98 and NT, IPLab for Windows performs image acquisition, enhancement, visualization, and analysis. Version 2.3 contains a series of functions to enhance control of, and interaction with, cameras and microscopes. The new 3D Multispectral Acquire command gives a unified way to control cameras, filter wheels, shutters and focus motors to grab images in three dimensions at multiple wavelengths. This command requires the Shutters & Filters extension and a supported hardware device. Users can plug in most popular cameras and microscope control equipment. Applications for which the software is suited include FISH (fluorescence *in situ* hybridization), GFP (green fluorescent protein), and live whole-cell studies.

Reader Service No. 110

EZ-2000

From Nikon Europe www.nikon.co.jp

Software for confocal microscopy

Released in January, Nikon's latest software package for the PCM2000 high resolution confocal microscope, EZ-2000, adds functionality to this bench-top system. Focusing on 2, 2.5 and 3D confocal morphological mapping the EZ2000 software includes image acquisition features such as simple image analysis, 3-D stacking, 3-D volume rendering, pixel integration, three channel live side-by-side display, panning and x-y scanning. It supports all popular file formats such as TIF, BMP, PIC and AV. The EZ-2000 software allows users to fully rotate the image in any direction to give an "all round view" of the specimen.

Reader Service No. 111

Electrophoresis

ImageMaster VDS 3.0

From Hoefer Pharmacia Biotech

www.apbiotech.com

One-dimensional image analysis software

This enhancement to the Pharmacia Biotech ImageMaster range has been redesigned specifically for the 32-bit environment of Windows 95 and Windows NT 4.0. The

ADVERTISEMENT



A Super Spectrum of Possibilities

Pierce SuperSignal®
Chemiluminescent Substrates for all your protein research needs.

Western Blotting Applications

SuperSignal® West Pico Chemiluminescent Substrate*
Picogram-level sensitivity at an amazingly economical price
Product # 34080EL, 500 ml
(Formerly SuperSignal® Chemiluminescent Substrate for Western Blotting)

SuperSignal® West Dura Extended Duration Substrate*
24-hour signal duration and super sensitivity
Product # 34075EL, 100 ml
(Formerly SuperSignal® ULTRA Chemiluminescent Substrate)

SuperSignal® West Femto Maximum Sensitivity Substrate*
The most intense signal and the greatest sensitivity attainable with a chemiluminescent substrate
Product # 34095EL, 100 ml
(Formerly SuperSignal® BLAZE™ Chemiluminescent Substrate)

ELISA

SuperSignal® ELISA Pico Chemiluminescent Substrate*
Picogram-level detection for plate-based assays
Product # 37070EL, 100 ml
(Formerly SuperSignal® LBA for Luminometer-Based Applications)

SuperSignal® ELISA Femto Maximum Sensitivity Substrate*
A more sensitive substrate with enhanced low-end linearity
Product # 37075EL, 100 ml
(Formerly PowerSignal™ ELISA Chemiluminescent Substrate)
*Power Reading

Call for your **FREE** Substrates Selection Guide Poster!

Call 800-874-3723 for product information or 815-968-0747 for the name of your local distributor.



a Hoelter Life Science Company

3747 M. Meador Rd. • PO Box 117 • Rockford, IL 61135 USA • Fax 815-968-7216
E-mail: TL@perce.com • Internet: <http://www.perce.com> © Pierce Chemical Company, 1998.

READER ENQUIRY NO. 365

update is designed to keep the same ease of use found in version 2.0, but incorporates Internet communication from within the software, a customizable desktop publishing report generator, automatic iso-molecular weight calibration through multiple standard lanes, dot/slot blot screening analysis and display, and automatic graphic gel reports for immediate publication.

Reader Service No. 112

AlphaMatch 2D

From Alpha Innotech www.alphainnotech.com

Two-dimensional gel analysis

AlphaMatch 2D is a software package for the PC designed to perform detailed analysis of 2D electrophoresis gels saved as TIFF images from a variety of imaging systems. An optional database package, 2D Database, provides a combination of image analysis and database for the comparison of spots expressed in 2D gels.

Reader Service No. 113

BioImage 2D Analyzer

From Genomic Solutions Inc.

www.genomicsolutions.com

Two-dimensional gel analysis

The main improvement in the latest version of the BioImage 2D Analyzer software is greater efficiency, hence speed, in the handling of gel images. The software automates the analysis of complex 2-dimensional gel separations. Other improvements include the capability of creating webmaps and more powerful database queries. The software analyses gel images from various sources such as flatbed scanners, CCD cameras and phosphorimagers. The spots on the gel are quantified using boundary detection algorithms, and pixel integration within those boundaries.

Reader Service No. 114

Application development

Imaq Vision Builder

From National Instruments www.natinst.com

Developing vision applications

Imaq Vision builder is a software tool that aids in the development of machine vision and scientific imaging applications. Developers work within an interactive vision development environment complete with a full set of machine vision and image processing functions. The program also takes a step towards simplifying vision application development by generating an application recipe containing the function calls and parameters needed by developers. Working from this application recipe, they can then build and optimize their application for speed using the software in LabWindows/CVI, LabVIEW, BridgeVIEW, or in ActiveX containers, such



Spot the whistle: outer (left) and internal (centre) features combined with the aid of C ImTcl.

as Microsoft Visual Basic and Visual C.

Reader Service No. 115

C_ImTcl

From Foster Findlay Associates

www.demon.co.uk/ffaltld

Tcl scripting language wrapper for the image processing and analysis library, C_Images 2D

C_Images 2D is a library of over 370 functions in C providing a prototyping tool for developing image processing and analysis applications. Functions range from high-level user interaction with the image through standard image processing to low-

level code allowing pixel-by-pixel access to any part of the image. FFA's new product, C_ImTcl, gives the scripting language full access to C_Images 2D, allowing the user to write scripts to build specialized applications. A demo of C_ImTcl can be downloaded from FFA's web-site. The screen shot above shows a captured video image and a radiograph from the British Museum of a Moche culture (Peruvian AD 100-700) ceramic whistling vessel. The C_ImTcl library was used to develop an application to align and superimpose the external and internal structures.

Reader Service No. 116

ADVERTISEMENT

Fluorescent HRP Substrates

Turn Blu at the Possibilities

Pierce QuantaBlu™
Chemifluorescent Substrates

Fluorogenic HRP substrates for plate-based applications

QuantaBlu™ Fluorogenic Peroxidase Substrate
Ideal for use with peroxidase enzymes
Product # 15169EL, 250 ml

QuantaBlu™ NS/K
Specifically designed for nonstop and kinetic assays
Product # 15162EL, 250 ml

Features/Benefits:

- More sensitive than TMB, OPD or ABTS Substrates
- Flexible – stopped, nonstopped or kinetic assays possible
- Large dynamic range (4 log peroxidase concentration range)
- Excellent 24-hour working solution stability
- Large Stokes shift: excitation/emission maxima of 325/420; range of 315-345/370-460
- Does not photobleach

Call for your FREE Substrates Selection Guide Poster!
Call 800-374-3723 for product information or
815-968-8747 for the name of your local distributor.

PIERCE
a PerkinElmer Company

2747 N. Meridian Rd. • PO Box 117 • Rockford, IL 61105 U.S.A. • Fax: (815) 968-7316
E-mail: TW@perkinel.com • Internet: <http://www.pierce.com> © Pierce Chemical Company, 1998

READER ENQUIRY NO.

Sequence analysis

MacVector 6.5

From Oxford Molecular Group

www.oxmol.com

Sequence analysis for the Mac

MacVector 6.5 contains several enhancements including user-defined primer testing, expanded protein tools and powerful graphics. Clustal W multiple sequence alignments now include dendrogram output and can be printed in colour. MacVector files can be exported in GCG format for use with the Wisconsin Package or read by Oxford Molecular's OMIGA for Windows.

Reader Service No. 117

Lasergene99

From DNASTAR

www.dnastar.com

Deliveries of this new software package begin in February 1999

Lasergene99, latest incarnation of the Lasergene suite of sequence analysis software tools, brings broad Internet connectivity to sequence analysis on Win95/98/NT and Macintosh computers. Lasergene99 supports Internet BLAST searching from all

analysis modules. And the GeneQuest gene discovery tool provides an Entrez query builder with direct access to NCBI Entrez for text searches. Lasergene99 Includes several convenience enhancements. There are explicit import/export options for a variety of public and proprietary data formats, and sequences from public databases may be entered into analysis modules simply by specifying accession numbers.

Reader Service No. 118

Spectroscopy

HyperUV/IR

From Shimadzu

www.shimadzu.com

Going by the book

Two new UV-Vis and FTIR software packages from Shimadzu contain the functions required for validation, IQ/OQ and GLP in accordance with the *European Pharmacopoeia*. HyperUV 1.50 and HyperIR 1.51 provide simple functions for the treatment and qualitative evaluation of spectra, as well as complex applications such as DIN H 18 analysis (oil in water) for FTIR. Functions for quantification, multi-component analysis and colorimetric measurements for UV-

Vis are also included, as well the usual functions for measurement and data processing.

Reader Service No. 119

Spectroscopy software

From Galactic Industries Corporation

www.galactic.com

A 12-page catalogue for the spectroscopist

This new catalogue features Galactic's line of spectroscopy software products. The catalogue includes detailed product descriptions and graphics on GRAMS/32, a data processing and data management package; Spectral ID, a spectral library searching tool that supports the popular commercial libraries; PLSplus/IQ, a spectroscopy chemometrics toolbox; and GRAMS/3D, a real-time 3D visualization package. System requirement details are also given for each product.

Reader Service No. 120

TD-2020 Luminometer

From Steptech

www.steptech.co.uk

Improved access to spreadsheets

The Turner Designs TD-20/20 research luminometer can measure light output using test tubes, culture dishes, and scintillation vials, as well as beads, strips, and other solids. The TD-20/20 is sensitive to 0.1 femtogram of luciferase and has a dynamic range of greater than 5 orders of magnitude, extendable to greater than 7 orders of magnitude with an optical filter. Spreadsheet Interface Software is now provided as standard, allowing the user to directly capture data from the luminometer to an Excel spreadsheet. This eliminates the need to transfer data from an ASCII file.

Reader Service No. 121

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENT



Pierce 1-Step™ Colorimetric Substrates

Ready-to-use, pre-formulated colorimetric substrates.

1-Step™ Substrates are available in the following formulations:

Slow TMB Product # 34024EL, 250 ml	NBT/BCIP Product # 34042EL, 250 ml
Turbo TMB Product # 34022EL, 250 ml	PNPP Product # 37621EL, 100 ml
ABTS Product # 37615EL, 250 ml	
Chloronaphthol Product # 34012EL, 250 ml	

Features/Benefits:

- One component
- Ready-to-use
- Stable

Call for your FREE Substrates Selection Guide Poster!
Call 800-874-3723 for product information or
815-966-0747 for the name of your local distributor.

PIERCE
a PerkinElmer Life Sciences Company

3747 N. Meridian Rd. • PO Box 117 • Rockford, IL 61125 U.S.A. • Fax: 815-966-7316
E-mail: 140@piercelife.com • Internet: http://www.piercelife.com © Pierce Chemical Company, 1998

READER ENQUIRY NO. 367

ADVERTISEMENT

**Detect residual DNA with Wako's
DNA Extractor Kit
or Isolate Genomic DNA with
DNA Extractor WB Kit**
Both use the Nal Method
No Phenol No Chloroform

Wako
Wako BioProducts
Wako Chemicals USA, Inc.
1600 Bellwood Road
Richmond, VA 23237
Telephone: (804) 271-7677
Facsimile: (804) 271-7791
(800) 992-9256
bioproducts@wakousa.com

READER ENQUIRY NO. 39